JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 8 (2020), Issue 9 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Corrigenda and Addenda

Viewpoint

A QR Code–Based Contact Tracing Framework for Sustainable Containment of COVID-19: Evaluation of an Approach to Assist the Return to Normal Activity

Ichiro Nakamoto[1], PhD; Sheng Wang[1], PhD; Yan Guo[1], PhD; Weiqing Zhuang[1], PhD

School of Internet Economics and Business, Fujian University of Technology, Fuzhou, China

Corresponding Author:
Sheng Wang, PhD
School of Internet Economics and Business
Fujian University of Technology
999 Dongsanhuang Road
JinAn District
Fuzhou
China
Phone: 86 132 550 66365
Fax: 86 0591 83503931
Email: fw107@foxmail.com

Abstract

We discuss a pandemic management framework using symptom-based quick response (QR) codes to contain the spread of COVID-19. In this approach, symptom-based QR health codes are issued by public health authorities. The codes do not retrieve the location data of the users; instead, two different colors are displayed to differentiate the health status of individuals. The QR codes are officially regarded as electronic certificates of individuals' health status, and can be used for contact tracing, exposure risk self-triage, self-update of health status, health care appointments, and contact-free psychiatric consultations. This approach can be effectively deployed as a uniform platform interconnecting a variety of responders (eg, individuals, institutions, and public authorities) who are affected by the pandemic, which minimizes the errors of manual operation and the costs of fragmented coordination. At the same time, this approach enhances the promptness, interoperability, credibility, and traceability of containment measures. The proposed approach not only provides a supplemental mechanism for manual control measures but also addresses the partial failures of pandemic management tools in the abovementioned facets. The QR tool has been formally deployed in Fujian, a province located in southeast China that has a population of nearly 40 million people. All individuals aged ≥3 years were officially requested to present their QR code during daily public activities, such as when using public transportation systems, working at institutions, and entering or exiting schools. The deployment of this approach has achieved sizeable containment effects and played remarkable roles in shifting the negative gross domestic product (–6.8%) to a positive value by July 2020. The number of cumulative patients with COVID-19 in this setting was confined to 363, of whom 361 had recovered (recovery rate 99.4%) as of July 12, 2020. A simulation showed that if only partial measures of the framework were followed, the number of cumulative cases of COVID-19 could potentially increase ten-fold. This approach can serve as a reliable solution to counteract the emergency of a public health crisis; as a routine tool to enhance the level of public health; to accelerate the recovery of social activities; to assist decision making for policy makers; and as a sustainable measure that enables scalability.

(JMIR Mhealth Uhealth 2020;8(9):e22321) doi:10.2196/22321

KEYWORDS

COVID-19; coronavirus; symptom-based; quick response; eHealth; digital health; telesurveillance; pandemic; epidemic; interoperability

Introduction

COVID-19, which is caused by SARS-CoV-2, has resulted in millions of confirmed cases and hundreds of thousands of deaths worldwide since the first cases were officially reported in December 2019; it has been declared a global pandemic by the World Health Organization (WHO) [1]. The negative impacts of COVID-19 are unprecedented and pervasive at a variety of levels, extensively spreading from individuals and institutions to public authorities. For individuals, COVID-19 is not only a

crisis of physical health but has substantially affected mental health, as evidenced by the growing number of consultations regarding psychological stress, anxiety, and depression [2]. Prior research shows that the percentages of the population who are concerned about symptoms, prevention and therapy, and psychological problems during the COVID-19 epidemic are 65%, 15%, and 11%, respectively; these levels are commensurate with those of the past outbreak of severe acute respiratory syndrome (SARS) [3,4]. Further, the pandemic has caused remarkable decreases in the gross domestic product (GDP) in many economies; thus, firms must employ measures to resume production and survive unexpected disasters in the future [5]. In the absence of credible and traceable information, it is difficult for governments to make coherent and accurate decisions [6]. Lifting the lockdown of COVID-19 is an enormous challenge because in the absence of effective screening and isolation aided by digital health, the second wave outbreak of COVID-19 cannot be precisely predicted contingent on present scientific understanding [5,7]. Subclinical infection through asymptomatic and presymptomatic transmission unnoticeable to individuals in close contact and delays in sharing of knowledge complicate the trajectory of the outbreak; therefore, the pandemic will take a longer time to subside [4,8]. The curve of the first wave of the outbreak may flatten, followed by a sizeable reduction in new infections, which will promote the opportunity to reopen the economy [8,9]. However, even if workplaces resume their interrupted production, doubts will be cast over whether it is feasible to sustain the same pace that was employed prior to the outbreak. Given the infectiousness of COVID-19 and the dynamics of high subclinical transmission, controlling the epidemic by mere manual contact tracing is less effective and infeasible. The WHO has forecast that the pandemic is still in its early stage, which necessitates long-term efforts in combating COVID-19. Hence, solutions to these challenges are of concern to individuals, firms, and public authorities that are struggling to return to normal rhythms [10].

A range of digital health approaches have been employed to contain the spread of disease during the current COVID-19 pandemic and past pandemics [11-20]. These control measures have been proved to be effective for numerous countries in productively depleting the first wave of COVID-19; among these measures, contact tracing is considered to be the centerpiece of containment, and it attracts a great deal of attention. Contact tracing involves identifying, quarantining, and alerting contacts of infected individuals [13]. Some countries that ceased contact tracing due to high prevalence of COVID-19 are currently reinstituting this strategy to curtail the potential impact of the second wave of the outbreak [11]. However, contentious issues facing contact tracing apps include the debate on which deployment framework (ie, centralized versus decentralized) and sensor technology (ie, Bluetooth, GPS, or quick response [QR]) can better address critical challenges such as effectiveness and sustainability of containment. Centralized architecture mostly complies with the Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) protocol, and personal data is collected for public control. In contrast, the decentralized approach mainly follows the Decentralized Privacy-Preserving Proximity Tracing (DP-3T) protocol; consequently, private information is not collected, and inference of exposure is constrained only to local devices [16]. However, even in the decentralized framework, a centralized database that at least accommodates the infected is of necessity to guarantee reliability. Hence, essentially, the concept of decentralization is only partially tenable [14-16]. Decentralized systems mostly do not retrieve sensitive data of users such as locations, and sharing permission is requested to preserve the users' privacy. These systems hinge substantially on peer-to-peer self-reporting that is trustworthy by the peer network; thus, high participation of the population is essential for meaningful containment. This mechanism may not completely address all likely routes of transmission between individuals even with full participation, accounting for potential delay in knowledge sharing [16]. The decentralized Bluetooth framework recently coprovided by Apple and Google supports preservation of privacy and provides users with anonymized contact exposure and guidance on how to respond if a risk is identified. However, because the nonidentifiable data are either changed or deleted periodically, traceability of the outbreak is jeopardized. Second, the definition of close contact is rather flexible; therefore, the detection precision of exposure may differ in different settings. Third, globally, only approximately one-quarter of smartphones are compatible with the Bluetooth standard required by Google and Apple, which inevitably inhibits the effectiveness of contact tracing [14-16]. The latest rebounding of confirmed cases in Japan has demonstrated the fragility of the Bluetooth framework [1]. On the other hand, extant contact tracing apps that collect GPS data to identify exposure improve the credibility of data. This supports evidence-based decision making but raises other concerns. The first concern is that location is a close proxy for contact but is not equal to it; therefore, the contact inference based on GPS data could generate confusion for the public due to bias in technical precision. The second concern is how to curb misuse and unauthorized access to sensitive information [16,19].

Beginning with contact tracing, telesurveillance has become more sophisticated, covering technologies from telecare to triage (eg, sorting of patients). Although the aforementioned tools and techniques may remarkably enhance the current capabilities of containment and reduce the spread of COVID-19, a broader framework is needed to increase its effectiveness through the integration of existing digital health measures [12,13,16,17]. The negative impacts of COVID-19 on physical health, mental health, and social rhythms are dramatic and far-reaching; hence, individuals, institutions, and public authorities urgently need integrated guidance and delay-free information that can safeguard the well-being of the public and assist them in making a smooth transition back to work and normal activities [18]. The lack of an inclusive repository with a uniform structure causes repeated failures for extant telesurveillance. Compared with the GPS-based approach, symptom-based QR contact tracing does not identify the location details of users; thus, the vulnerability inherent to the GPS-based approach can be waived. Further, credibility and traceability can be enhanced relative to the self-report mechanism used in the Bluetooth approach [16,18,19].

In contrast to location-coupled QR contact tracing, here, we evaluate a symptom-based QR framework that addresses the

disadvantages of GPS and Bluetooth. The major aim of this paper is to appraise how this approach can serve as an urgent countermeasure against the COVID-19 crisis as well as a routine tool to guide the return of daily activities in workplaces, travel locations, and communities to normal rhythms.

Core Concept and Framework of the Tool

Features for Individuals

The first crucial concept used in the approach is symptom-based QR health codes issued by public health authorities. The codes do not retrieve the location data of the users; instead, two colors are supported to differentiate the health status of individuals. The QR codes are illustrated in Figure 1. A green code denotes that the individual is not infected with COVID-19 according to a polymerase chain reaction (PCR) test record; thus, the individual passes the health verification test. In contrast, an orange code signals that the person is either infected with COVID-19 or the likelihood of being infected is high. This can be stratified into six scenarios in which the individual:

1. Is infected with COVID-19
2. Had close contact with an individual who is infected with COVID-19
3. Comes from a region where the infection rate of COVID-19 is high
4. Is a resident of a community under strict public surveillance due to severe infection
5. Has a record of a high fever within the past 14 days
6. Has a record indicating the purchase of anti-fever medicine within the past 14 days

The QR codes are officially regarded as electronic certificates of individuals' health status. The information in the codes is automatically read and analyzed by QR scanners. The core notions of the design are to enhance the credibility of data, increase the speed of processing, and reduce errors arising from manual operation.

The second crucial concept is the synthesis of critical features, including contact tracing, exposure risk self-triage, self-update of health status, health care appointments, contact-free psychiatric consultation, and QR codes for other family members. The platform also coherently merges health insurance and prescription services, which reduces the costs and delays of operation for users.

Figure 1. Color schema of the quick response codes: (A) green; (B) orange.

Features for Institutions

The QR approach consolidates features that can assist companies and institutions with early and automatic case screening and flexible social distancing based on QR information. This avoids crosstransmission between workers and guarantees smooth resumption of production at the normal rhythms. Each worker can double-check the health colors of their colleagues (no other private information is shared), which lowers systematic mistakes of mis-surveillance. The accounts of the QR scanning sites are configurable and scalable. Firms can choose to report local statistics to public authorities for subsequent analysis. Coupled with proximity detection techniques, an alarm can be raised in scenarios of potential close contact. Managers can thus establish a flexible and correctable threshold to implement automatic reminders of social distancing as science gains updated knowledge of the disease.

Features for Public Authorities

The QR approach can help policy makers make more effective decisions and improve public surveillance. The accounts of QR scanning sites at travel spots are administered by public authorities. Any individual must present their QR code to use public transportation systems such as buses, railways, or airports. The summary at each site is synchronized with public authorities for general analysis.

The QR-based uniform design increases the seamless synergy among individuals, institutions, and public authorities and reduces delays in information processing and transmission. In Table 1, we illustrate the core concepts and potential scenarios of implementation of the QR approach. Figures 2 and 3 present a brief diagram and screenshots of the approach, respectively.

Table 1. Concepts of the QR-based framework.

Concept and subject	Description
Products	
Individuals	Mobile apps (iOS and Android)
Institutions and public authorities	Web-based platforms (Windows)
Scenarios	
Individuals	Electronic certificates of health status
	Dynamic of updates health status
	Self-triage of exposure risk and self-isolation
	Contact tracing of past potential exposure
	Noncontact health care appointments when infected or suspected of being infected
	QR[a] codes for other family members
	Remote consultations on psychiatric care
Institutions	Surveillance of the health status of employees and close contacts
	Setup of automatic social distancing alerts
	Mutual surveillance between employees
Public authorities	Issuance and administration of QR codes
	QR-based travel control
	QR code scanning account management
	Statistics and outbreak inference
Strengths	
Participation	Equity of access and high participation
Cost	Reduced costs of manual operation and cross-platform synergy
Errors	Reduced errors of manual operation
Other	Sustainability and scalability

[a]QR: quick response.

Figure 2. Diagram of the QR-based approach integrating features critical to individuals, institutions, and public authorities. QR: quick response.

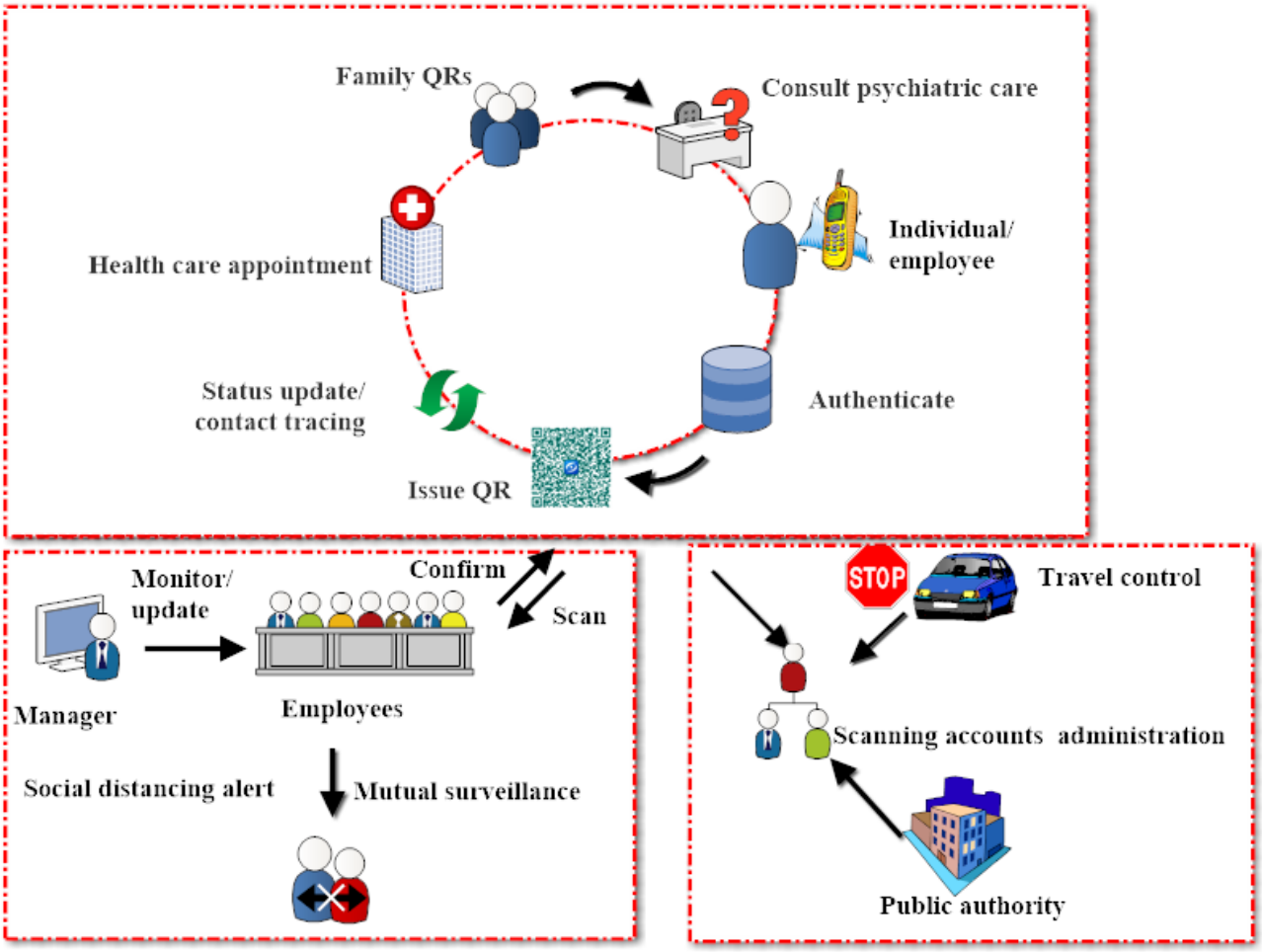
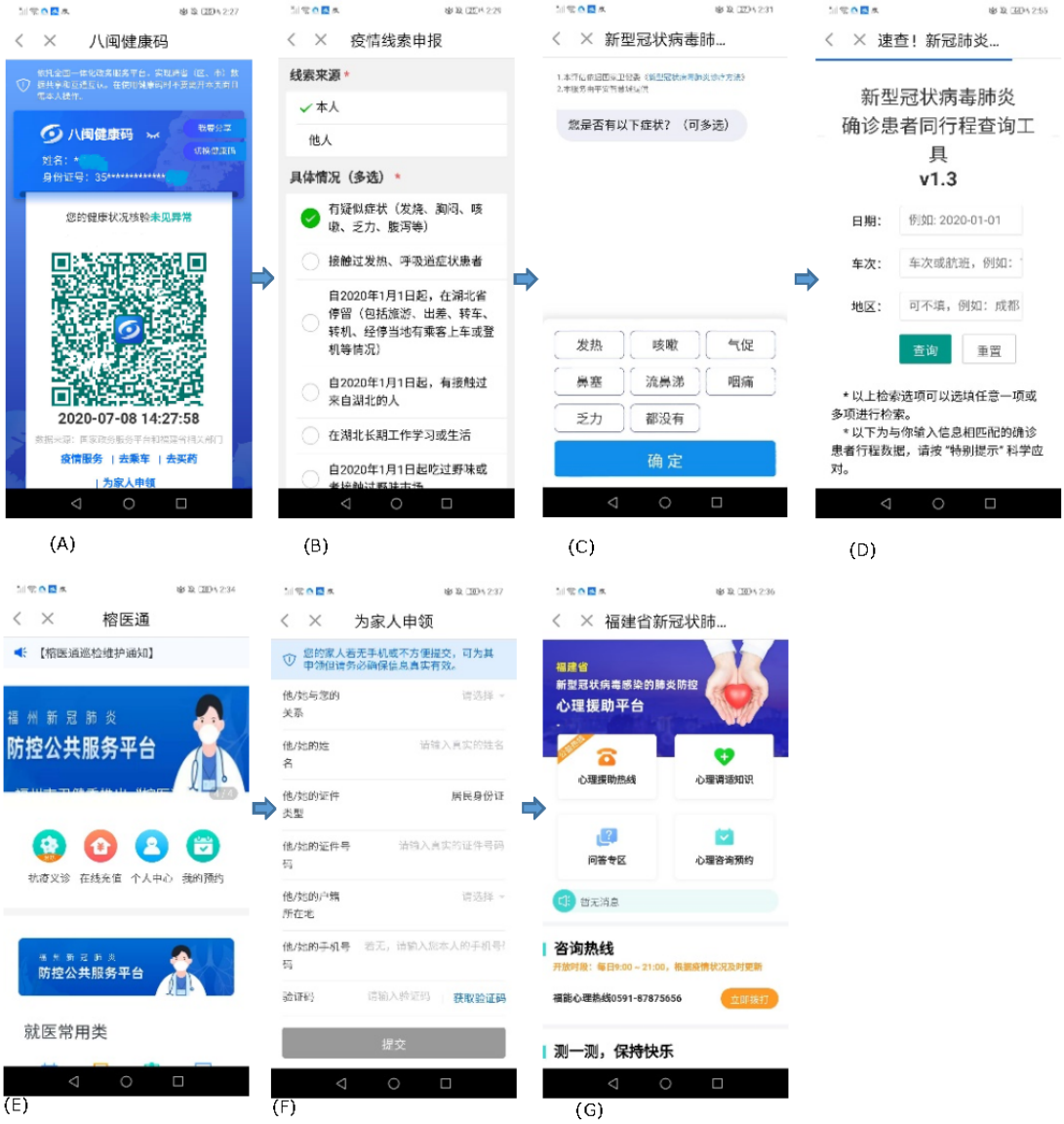


Figure 3. Screenshots of the QR-based approach for individuals: (A) QR code, (B) health status update, (C) risk self-triage, (D) contact tracing, (E) health care appointment, (F) request for QR codes for family members, (G) psychiatric consultation.



Quantitative Analysis of the Effectiveness of the Approach

To apply a cross-platform surveillance approach and to determine the best way to integrate discrete components into a uniform health system, evidence is required of the effectiveness of the approach in practice, with particular scrutiny of the context in which it is to be deployed [3].

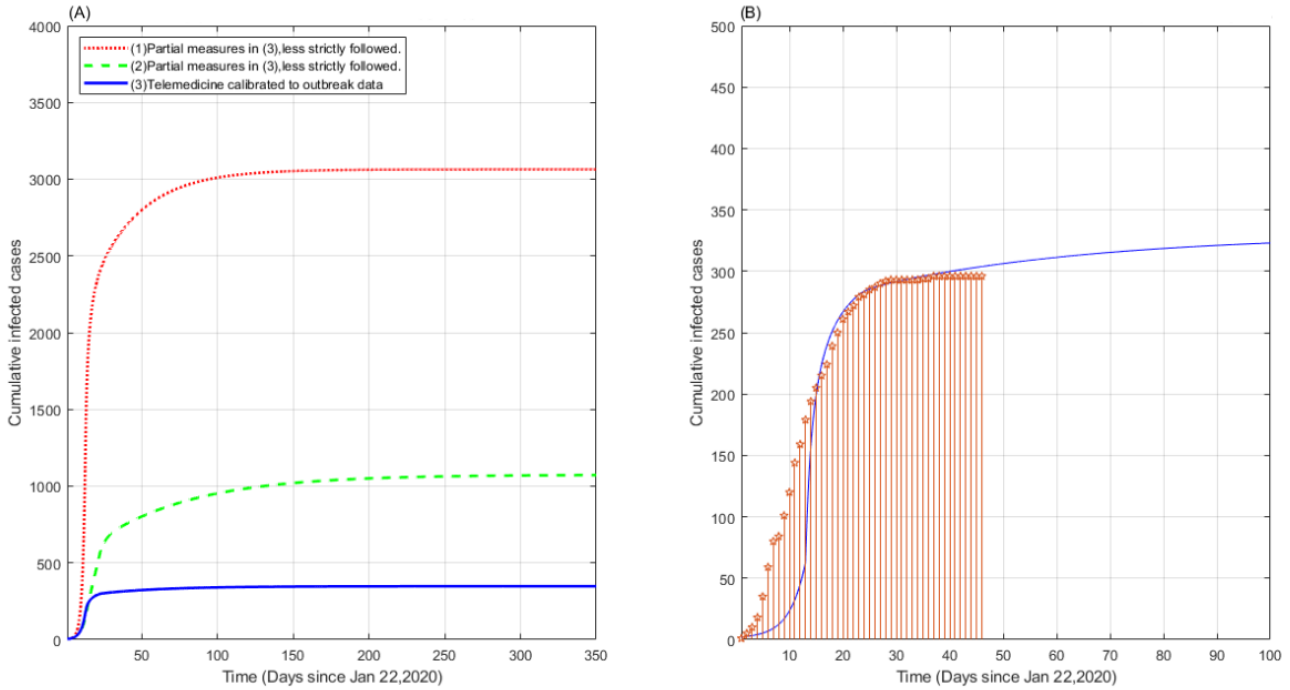
Since the spread of COVID-19 began, the QR tool has been formally deployed in Fujian, a province located in southeast China that had a population of nearly 40 million people by late June 2020. All individuals aged ≥3 years were officially requested to present their QR code during daily public activities, such as when using public transportation systems, working at institutions, and entering or exiting schools. Almost 5 million individuals left Wuhan, the epicenter of the COVID-19 pandemic in China, when the outbreak occurred. Approximately one-third of these individuals traveled to other regions, such as Fujian Province [21]. With the app in use, by July 12, 2020, the cumulative positive cases of COVID-19 were confined to 363 individuals, of which 361 (99.4%) recovered from the disease. Only one patient died and one patient remained positive for COVID-19 at that point. Meaningfully effective containment was achieved through a strict strategy of centralized control and extensive deployment of the tool. The latest data show that since the outbreak started, the GDP of provinces in China, including Fujian, decreased to –6.8% in the first quarter of 2020 due to the nationwide lockdown. However, due to the gradual lifting of lockdown and reopening of production, the GDP bounced back and became positive by July [5].

As shown in Figure 4, we deployed the model introduced in [22] to simulate the heterogeneous epidemic evolution of the COVID-19 outbreak in Fujian using data published by Johns Hopkins University [23]. The epidemiological mean-field framework sketched in [22] captures the average effect involving the whole population, which extends the classical susceptible-infected- recovered (SIR) model and evaluates more complicated disease transmission scenarios partitioned into

eight stages of infection, including susceptible, infected, diagnosed, ailing, recognized, threatened, healed, and died. The status of each stage is determined by the interactions among different adjacent stages. The core concept hinges on the deliberation that subdivided models enhance the accuracy of portraying the dynamic spread of COVID-19. This model estimates how progressive countermeasures would affect the spread of the pandemic. The system has been shown to correctly delineate the dynamics of the epidemic and is suitable for the prediction of containment measures with varying strengths and natures. The simulation denotes that restrictive social distancing measures must be effectively combined with contact tracing and other countermeasures to decrease the spread of COVID-19. The assumptions of the model are that the stages of infection are mutually exclusive; the likelihood of being susceptible again after recovering from the infection is negligible; there is a distinction between undiagnosed and diagnosed individuals; and there is a delay in the emergence of symptoms.

We calibrated the model to the Fujian outbreak data starting from January 22, 2020, when the first case was identified there (see Figure 4(A), blue solid curve, and Figure 4(B)). The curves in Figure 4(A) denote the cases where the features of digital health in our framework were adopted in chronological order to counteract the spread of disease up to a 1-year horizon. The imposed countermeasures are as follows: (1) social distancing, including in workplaces, and mutual surveillance between workers is followed from day 2; (2) contact tracing of individuals in close contact with infected people is appended starting on day 12; (3) contact tracing of individuals with a record of high fever or record of the purchase of anti-fever medicine is added on day 22; (4) contact-free psychiatric consultations and travel control measures are followed. Comparatively, these measures can be employed simultaneously in practice. The blue solid curve (Figure 4(A)) corresponds to the approach using all the above four measures. The simulated cumulative cases amounted to nearly 350 cases by July 2020, which is close to the observed outbreak data in practice. The red dotted curve indicates the cases where (1) and (2) are followed but less strictly, after which the number of cumulative cases would increase almost ten-fold to 3100 half a year later. In contrast, the green dashed curve implies the case where measures (2), (3), and (4) are more loosely followed. This scenario caused about three-fold growth in the number of cumulative cases to 1100 cases by July. Hence, the strict deployment of the merged measures remarkably contained the spread of COVID-19.

Figure 4. Simulation of heterogeneous countermeasures during the COVID-19 outbreak. (A) Model simulation of different scenarios; (B) comparison of the model simulation with the observed data.



Strengths and Limitations of the Approach

The centralized approach is distinct from other decentralized models in three paramount ways. The first is the integration of features that are important to the concerned population in one identical platform; this is pivotal when combating highly contagious diseases, as delays of data sharing can enable increased spread of COVID-19. Tools used at the individual or public authority level may not respond in a timely and accurate manner to requirements at the institutional level. The collective provision of features oriented toward a wider range of users promotes broader applicability of the platform. In lieu of the spontaneous self-reporting that is mostly used in decentralized frameworks, where the agreement of individuals is requested, this approach affirms high, wait-free participation. As delays and misinformation are still of public concern, this responds to the urgent need of a uniform framework internalizing various sources of corroboration and efforts. Second, the tool facilitates self-triage for individuals and self-scheduling for institutions. Hence, load balancing of the pressure on overburdened health care systems is plausible. The tool also prevents unnecessary

and risky in-person visits. Third, due to its interoperability, credibility, and traceability, the QR approach can play valuable roles as both an emergent and routine tool to counteract COVID-19.

Several concerns must be addressed to ensure successful implementation of this tool. First, it is important to prevent malicious or unauthorized use of the QR data. Second, for economic settings where the general population is highly sensitive to preservation of privacy, acceptance of this approach will be compromised; thus, the effects of containment must be closely observed.

Illegal or unauthorized use of health care information is detrimental at a variety of levels [24]. Reports forecast that misuse of this information will cause an average financial loss of nearly US $7.13 million worldwide in 2020 [9]. Unlawful use of data can damage the reputation of service providers and negatively impact the confidence and health of patients. The Harris study targeting 1527 individuals found that 123/1527 (8.1%) of the respondents refrained from participating in health programs due to reduced confidence in the quality of data protection [25]. Another study conducted with 3959 individuals suggested that unauthorized use of patients' information would increase the likelihood of negative perceptions and responses. It was found that almost two-thirds of the participants (2764/3959, 69.8%) expressed security-related concerns, which impeded subsequent health care interventions and interactions [26]. A simulation performed at Oxford University implied that digital contact tracing interventions would lose the potential to substantially contain the spread of COVID-19 if the population uptake rate decreased approximately below 60%, even if these interventions were deliberately implemented alongside other countermeasures [27].

Conclusions

The framework presented in this study has substantially controlled the spread of COVID-19 in Fujian Province, where almost 40 million people reside, since the first case was officially reported in January 2020. Of the 363 reported cumulative cases, 361 (99.4%), recovered, 1 patient died (mortality rate 0.3%) and one patient remained positive for COVID-19 (0.3%) as of July 12, 2020. Due to the early deployment and rigid implementation of the approach, it successfully helped transition the GDP from –6.8% to a positive value by July. Firms have succeeded in gradually resuming normal production without sacrificing effective containment. The tool has assisted public authorities with effective containment and travel control. The model simulation shows that if the partial measures were followed less rigorously, the likelihood of confirmed cases would increase, leading to multiple times of growth in the number of cumulative cases of COVID-19. Due to the credibility, interoperability, and sustainability empowered by the design, long-term containment of COVID-19 is feasible. Digital health is a principal factor contributing to the success of containment; however, it cannot solve all the challenges the world is presently facing [3]. The convergence of COVID-19 diagnostics and treatment provides opportunities to deliver potentially disruptive technologies to drive the development of integrated health systems. These systems should increase accessibility to contain the pandemic while improving the promptness, interoperability, and credibility of outbreak detection and surveillance while guiding more precise and sustainable public health responses [28].

The deployment of telesurveillance to specific settings must account for both technical and nontechnical factors. The effectiveness of telesurveillance may vary due to cultural conflicts as well as users' moral and religious backgrounds in different countries and regions [29]. The collaboration of stratified participants should be guided by law enforcement for better protection of individuals' data, preventing malicious breaches of privacy information as well as abuse beyond the scope of legal screening, contact tracing, and surveillance [30]. This symptom-based QR approach facilitates the optimized allocation of limited health care resources [28,31]. It clearly aids the identification and isolation of cases at an earlier stage and generates seamless, delay-free cooperation of individuals, institutions, public authorities, and other responders in both the short term and in the long term. The approach can be used both to effectively counteract the emergency of a public health crisis and as a routine surveillance technique in the postpandemic era to facilitate rapid recovery from the shock of the outbreak. The integration of features that are critical for the containment of COVID-19 in a uniform platform will be a research trend to achieve more effective control. The capabilities of rapid response, traceability, and credibility provided by this approach can help society to achieve a balance between sustainable containment and smooth recovery of the economy. This tool is scalable for extension of functionality with advances in artificial intelligence, big data, and other technologies. It will enable coordinated data-sharing mechanisms ahead of, during, and after an epidemic, improving the quality and sustainability of data in an unprecedented era of high-impact and cyclical pandemics. Information obtained from the app can also increase scientific understanding of the dynamics of COVID-19 and deliver positive insight for other infected communities.

Acknowledgments

This research is partly funded by the FuJian Philosophy and Social Sciences Association (No. FJ2019B081), Science and Technology Development Center of the Ministry of Education (No. 2019J01008), Key Projects of the National Social Science Foundation (No. 19AGL017), Ministry of Education's Humanities and Social Sciences Research Project (No. 18YJAZH153), Fujian Natural Science Foundation (No. 2018J01648), and Development Fund of Scientific Research from Fujian University of Technology (No. GY-S1810).

Authors' Contributions

IN designed the research and drafted the manuscript. SW analyzed the content. IN, SW, and WQZ contributed to the modification of the manuscript draft. SW, WQZ, and YG provided framework analyses of the content. All authors read and approved all sections of the final manuscript for submission.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) Situation Report – 168. World Health Organization. 2020 Jul 06. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200706-covid-19-sitrep-168.pdf?sfvrsn=7fed5c0b_2 [accessed 2020-07-07]
2. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
3. Gao Y, Liu R, Zhou Q, Wang X, Huang L, Shi Q, COVID-19 Evidence and Recommendations Working Group. Application of telemedicine during the coronavirus disease epidemics: a rapid review and meta-analysis. Ann Transl Med 2020 May;8(10):626 [FREE Full text] [doi: 10.21037/atm-20-3315] [Medline: 32566563]
4. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 2004 Oct 15;160(6):509-516 [FREE Full text] [doi: 10.1093/aje/kwh255] [Medline: 15353409]
5. Wang Y, Hong A, Li X, Gao J. Marketing innovations during a global crisis: A study of China firms' response to COVID-19. J Bus Res 2020 Aug;116:214-220 [FREE Full text] [doi: 10.1016/j.jbusres.2020.05.029] [Medline: 32501308]
6. Charette RN. Managing risk and responsibility during a crisis: The COVID-19 pandemic has resurfaced questions about the government's role when disaster strikes - [Spectral Lines]. IEEE Spectr 2020 May;57(5):4-4. [doi: 10.1109/mspec.2020.9078399]
7. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature 2005 Dec 17;438(7066):355-359 [FREE Full text] [doi: 10.1038/nature04153] [Medline: 16292310]
8. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One 2020;15(3):e0230405 [FREE Full text] [doi: 10.1371/journal.pone.0230405] [Medline: 32231374]
9. Cost of a Data Breach Report 2020. IBM. 2020. URL: https://www.ibm.com/security/data-breach [accessed 2020-08-12]
10. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020 May;5(5):e261-e270 [FREE Full text] [doi: 10.1016/S2468-2667(20)30073-6] [Medline: 32220655]
11. Breeher L, Boon A, Hainy C, Murad MH, Wittich C, Swift M. A Framework for Sustainable Contact Tracing and Exposure Investigation for Large Health Systems. Mayo Clin Proc 2020 Jul;95(7):1432-1444. [doi: 10.1016/j.mayocp.2020.05.008] [Medline: 32561146]
12. McCall B. Shut down and reboot—preparing to minimise infection in a post-COVID-19 era. Lancet Digit Health 2020 Jul;2(6):e293-e294 [FREE Full text] [doi: 10.1016/S2589-7500(20)30103-5] [Medline: 32346671]
13. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491) [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
14. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR mHealth uHealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
15. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR mHealth uHealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
16. LI J, Guo X. Global Deployment Mappings and Challenges of Contact-tracing Apps for COVID-19. SSRN Journal 2020 May 24. [doi: 10.2139/ssrn.3609516]
17. Sacks JA, Zehe E, Redick C, Bah A, Cowger K, Camara M, et al. Introduction of Mobile Health Tools to Support Ebola Surveillance and Contact Tracing in Guinea. Glob Health Sci Pract 2015 Dec 12;3(4):646-659 [FREE Full text] [doi: 10.9745/GHSP-D-15-00207] [Medline: 26681710]
18. Ekong I, Chukwu E, Chukwu M. COVID-19 Mobile Positioning Data Contact Tracing and Patient Privacy Regulations: Exploratory Search of Global Response Strategies and the Use of Digital Tools in Nigeria. JMIR mHealth uHealth 2020 Apr 27;8(4):e19139 [FREE Full text] [doi: 10.2196/19139] [Medline: 32310817]
19. Wang S, Ding S, Xiong L. A New System for Surveillance and Digital Contact Tracing for COVID-19: Spatiotemporal Reporting Over Network and GPS. JMIR mHealth uHealth 2020 Jun 10;8(6):e19457 [FREE Full text] [doi: 10.2196/19457] [Medline: 32499212]

20. Scott BK, Miller GT, Fonda SJ, Yeaw RE, Gaudaen JC, Pavliscsak HH, et al. Advanced Digital Health Technologies for COVID-19 and Future Emergencies. Telemed J E Health 2020 May 26. [doi: 10.1089/tmj.2020.0140] [Medline: 32456560]
21. Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020 Mar;395(10226):764-766. [doi: 10.1016/s0140-6736(20)30421-9]
22. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020 Jun;26(6):855-860 [FREE Full text] [doi: 10.1038/s41591-020-0883-7] [Medline: 32322102]
23. Johns Hopkins University COVID-19 Data. GitHub. URL: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv [accessed 2020-07-04]
24. Guttieres D, Stewart S, Wolfrum J, Springs SL. Cyberbiosecurity in Advanced Manufacturing Models. Front Bioeng Biotechnol 2019;7:210 [FREE Full text] [doi: 10.3389/fbioe.2019.00210] [Medline: 31552236]
25. Institute of Medecine. In: Nass SJ, Levit LA, Gostin LO, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC: National Academies Press; 2009.
26. 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]
27. Digital contact tracing can slow or even stop coronavirus transmission and ease us out of lockdown. University of Oxford. URL: https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown [accessed 2020-08-12]
28. Rubin EJ, Baden LR, Morrissey S, Campion EW. Medical Journals and the 2019-nCoV Outbreak. N Engl J Med 2020 Feb 27;382(9):866. [doi: 10.1056/NEJMe2001329] [Medline: 31986242]
29. Hernandez-Orallo E, Manzoni P, Calafate CT, Cano J. Evaluating How Smartphone Contact Tracing Technology Can Reduce the Spread of Infectious Diseases: The Case of COVID-19. IEEE Access 2020;8:99083-99097. [doi: 10.1109/access.2020.2998042]
30. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. Nature 2019 Feb;566(7745):467-474 [FREE Full text] [doi: 10.1038/s41586-019-0956-2] [Medline: 30814711]
31. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]

Abbreviations

DP-3T: Decentralized Privacy-Preserving Proximity Tracing
GDP: gross domestic product
PCR: polymerase chain reaction
PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing
QR: quick response
SARS: severe acute respiratory syndrome
SIR: susceptible-infected-recovered
WHO: World Health Organization

Edited by L Buis; submitted 08.07.20; peer-reviewed by J Li; comments to author 10.07.20; revised version received 28.07.20; accepted 18.08.20; published 07.09.20.

Please cite as:
Nakamoto I, Wang S, Guo Y, Zhuang W
A QR Code–Based Contact Tracing Framework for Sustainable Containment of COVID-19: Evaluation of an Approach to Assist the Return to Normal Activity
JMIR Mhealth Uhealth 2020;8(9):e22321
URL: http://mhealth.jmir.org/2020/9/e22321/
doi:10.2196/22321
PMID:32841151

Viewpoint

Digital Media's Role in the COVID-19 Pandemic

Huanyu Bao[1], MSc; Bolin Cao[2], PhD; Yuan Xiong[1], MSc; Weiming Tang[1], MD, PhD

[1]UNC Project-China, Guangzhou, China

[2]Shenzhen University, Shenzhen, China

Corresponding Author:
Weiming Tang, MD, PhD
UNC Project-China
No. 2 Lujing Road
Guangzhou, 510095
China
Phone: 86 15920567132
Email: weimingtangscience@gmail.com

Abstract

The severe acute respiratory syndrome coronavirus 2 outbreak has had a significant impact on global health, the economy, and society as a whole. Various measures are being taken to respond to the pandemic, with digital media playing a pivotal role, especially in the use of visual data to disseminate information, mobile health to coordinate medical resources, social media to promote public health campaigns, and digital tools to assist population management and disease tracing. However, digital media also faces some challenges like misinformation, lack of guidance, and information leakage. We encourage the increased use of digital media with a focus on improving trust, building social solidarity, reducing chaos, educating the public on prevention measures, and reducing the medical burden in facility-based sites.

(JMIR Mhealth Uhealth 2020;8(9):e20156) doi:10.2196/20156

KEYWORDS

COVID-19; digital health; media; pandemic; public health; social media; dissemination; health information; mobile health

As of May 31, 2020, the coronavirus disease (COVID-19) outbreak has led to the death of 367,166 people [1]. The pandemic is causing severe damage to the health care system, the economy, and society as a whole. Plans and actions to prevent and respond to the pandemic are urgently needed. Of these, the digital media's response, advocate, and mobilization plays an essential role. With the development of information and technology, digital media plays a pivotal role in this pandemic, especially in the use of visual data to disseminate information, mobile health (mHealth) to coordinate medical resources, and social media to promote public health campaigns.

First, visual data is used increasingly to demonstrate the distribution, transmission, and trend of this coronavirus outbreak. The unprecedented pandemic has brought an enormous amount of real time data, and many online media platforms adopted visual graphs to release COVID-19 statistics, which were rarely used during the severe acute respiratory syndrome outbreak. Data visualization can help people easily and efficiently process a large volume of information on disease transmission to understand the patterns of epidemics [2]. An example of this is the interactive dashboard developed by Johns Hopkins University based on the crowdsourcing data [3]. This dashboard provides data-driven visuals (eg, global cases map, critical data trends, latest news, and COVID-19 basics) to illustrate the situations of pandemics around the world, enabling the public and researchers to understand and monitor the outbreak timely. Similarly, a popular messaging app (WeChat) in China offered a location-based feature, "Cases Nearby," to show the location of the confirmed cases around the users and the places the cases have been without disclosing any personal information [4]. This visual footprint keeps users informed of the outbreak and advises them to take targeted measures to avoid high-risk areas. In addition, in the Prince of Wales Hospital of Hong Kong, an infographic on the principles of airway management was developed in 17 languages and disseminated through online social network platforms, benefiting other medical units to incorporate infection control procedures to reduce the transmission of COVID-19 [5].

Second, mHealth is surging in demand to reduce the overloading of health care systems. To avoid the high risk of contact with infected individuals, several virtual teleconsultation platforms (*EmergencyEye* in Germany, *Vodacom* in South Africa, and *WeDoctor* in China) were used to assist health care professionals. Facebook groups have been used by doctors to share and integrate experiences in disease treatment and research in real time, a subgroup called the PMG COVID19 has 36,900

members worldwide [6]. The pandemic has also driven research and the application of artificial intelligence (AI) in dealing with this emerging issue. By using lung computed tomography scans, AI technology was used to help doctors make a quick judgment of coronavirus pneumonia [7]. To help fight mental health disorders during the pandemic [8], an AI-based chatbot has also played important roles in responding to people's emotions and providing online consultation. This trend was witnessed by the surge in users for some Indian chatbot software during the outbreak [9].

Third, social media platforms are applied to educate people to take public health measures. As one of the first countries hit by COVID-19, Singapore's successful response has benefited from the early action taken by the country via social media. A national WhatsApp channel was immediately created to inform people living in Singapore about government updates and initiatives on COVID-19. There have been over 635,000 people subscribed to the channel to receive updated messages [10]. In China, the government has partnered with mobile phone operators to send automated text messages at various times throughout the day to keep people informed and alert them to keep a social distance. Additionally, in partnership with the health ministry, a Vietnamese music artist wrote a song, and a local dancer choreographed a dance on how to wash hands carefully and started a dance challenge on TikTok (a popular video-sharing app) [11]. The dance challenge video has gone viral and invited millions of people to learn about the essential steps of handwashing, playing a critical role in fighting against the spread of COVID-19.

Fourth, digital tools are applied to assist the management of work resumption and citizen migration after the pandemic. In China, the digital health code, which displays a Quick Response (QR) code with an individual's health status, is widely used to track citizen's health status and estimate their potential risk in transmitting the virus. Individuals are assigned a color code—green, yellow, or red—that indicates their health status. The functions of digital health codes are two-fold: for ensuring anyone entering a public place is healthy and for contact tracing purposes. Although such digital tools have raised concerns about privacy, it helps to contain the outbreak of epidemics and mitigate the burden of public health surveillance, allowing society to return to normal. In addition, coronavirus tracking apps were also applied with the official government in other countries to a id contact tracing, such as Australia (*COVIDSafe*), Bahrain (*BeAware Bahrain*), Colombia (*CoronApp*), and Ghana (*GH Covid-19 Tracker App*).

Although digital media has made considerable efforts in response to the pandemic, it is still facing some challenges. First, misinformation is a pressing problem. Rumors, fake news, and deliberate misinformation have been spreading on social media platforms, causing distrust and further endangering public health [12,13]. To respond to the infodemic caused by misinformation, some efforts have been made to correct the misinformation. Useful corrective actions such as more coherent information that provides alternative explanations to misleading information and appeals to credibility should be continuously, widely, and frequently distributed [14]. Second, there is a lack of formalized guidance to guide the use of digital media in large-scale epidemics. In particular, personal privacy and data leakage is an issue that needs to be addressed urgently. *Zoom*, the teleconferencing software that was heavily used in this outbreak, suffered several hacks [15], which raised concerns about the security of digital tools. Although the digital health code has been used in several countries, there is still a concern about privacy, and personal information leakage may cause geographical discrimination against people from high-risk areas.

The COVID-19 pandemic is continuing to worsen, and more effective strategies are needed. Although digital media has played an important role, we strongly recommend that it should be further used to improve trust, build social solidarity, reduce chaos, educate the public for prevention measures, and reduce the medical burden in facility-based sites. Only by using multiple resources and working together globally, can we mitigate the effects of COVID-19, even if this comes at a cost.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2017YFE0103800), the National Nature Science Foundation of China (81903371), National Institute of Mental Health (R34MH119963), the National Science and Technology Major Project (2018ZX10101-001-001-003), and the Special Found on Prevention and Control of New Coronary Pneumonia in Guangdong Universities (2020KZDZX1047).

Conflicts of Interest

None declared.

References

1. World Health Organization. WHO coronavirus disease (COVID-19) dashboard URL: https://covid19.who.int/
2. Heer J, Bostock M, Ogievetsky V. A tour through the visualization zoo. Commun ACM 2010 Jun;53(6):59-67. [doi: 10.1145/1743546.1743567]
3. Rajgor DD, Lee MH, Archuleta S, Bagdasarian N, Quek SC. The many estimates of the COVID-19 case fatality rate. Lancet Infect Dis 2020 Mar. [doi: 10.1016/s1473-3099(20)30244-9]

4. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/S0140-6736(20)30260-9]
5. Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Mar 30 [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]
6. Smith M, Cortez MF. Bloomburg. 2020 Mar 24. Doctors turn to social media to develop Covid-19 treatments in real time URL: https://www.bloomberg.com/news/articles/2020-03-24/covid-19-mysteries-yield-to-doctors-new-weapon-crowd-sourcing [accessed 2020-04-25]
7. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology 2020 Mar 19 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
8. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
9. Preetipadma. Analytics Insight. 2020 Apr 13. Top 6 chatbots helping to curb COVID-19 induced mental anxiety URL: https://www.analyticsinsight.net/top-6-chatbots-helping-curb-covid-19-induced-mental-anxiety/ [accessed 2020-04-26]
10. UNDP. 2020. Innovations in Singapore's COVID-19 response URL: https://sgtechcentre.undp.org/content/sgtechcentre/en/home/blogs/covid19response.html
11. O'Kane C. CBS News. 2020 Mar 04. Catchy PSA about coronavirus turns into viral TikTok challenge about washing your hands URL: https://www.cbsnews.com/news/coronavirus-prevention-vietnam-song-ghen-co-vy-tik-tok-challenge-about-washing-hands-quang-dang-john-oliver/ [accessed 2020-04-25]
12. Limaye RJ, Sauer M, Ali J, Bernstein J, Wahl B, Barnhill A, et al. Building trust while influencing online COVID-19 content in the social media world. Lancet Digital Health 2020 Jun;2(6):e277-e278. [doi: 10.1016/s2589-7500(20)30084-4]
13. Pennycook G, McPhetres J, Zhang Y, Rand D. Fighting COVID-19 misinformation on social mediaxperimental evidence for a scalable accuracy nudge intervention. PsyArXiv 2020 Mar 17. [doi: 10.31234/osf.io/uhbk9]
14. Walter N, Murphey ST. National Communication Association. 2018 Jun 14. Does correcting misinformation really work? URL: https://www.natcom.org/communication-currents/does-correcting-misinformation-really-work [accessed 2020-05-30]
15. Zoom. Support during the COVID-19 pandemic URL: https://zoom.us/docs/en-us/covid19.html

Abbreviations

AI: artificial intelligence
COVID-19: coronavirus disease
mHealth: mobile health

Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by W Cheng, J Jiang, E Da Silva, K Wall; comments to author 25.05.20; revised version received 01.06.20; accepted 11.06.20; published 18.09.20.

Please cite as:
Bao H, Cao B, Xiong Y, Tang W
Digital Media's Role in the COVID-19 Pandemic
JMIR Mhealth Uhealth 2020;8(9):e20156
URL: https://mhealth.jmir.org/2020/9/e20156
doi:10.2196/20156
PMID:32530817

Review

Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review

Daniel Fuller[1,2,3*], PhD; Emily Colwell[1*], MSc; Jonathan Low[1*], MSc; Kassia Orychock[1], MSc; Melissa Ann Tobin[4], MSc; Bo Simango[1], MSc; Richard Buote[4], MSc; Desiree Van Heerden[5]; Hui Luan[6], PhD; Kimberley Cullen[1,3], PhD; Logan Slade[4], BSc; Nathan G A Taylor[7], MPH

[1]School of Human Kinetics and Recreation, Memorial University, St. John's, NL, Canada

[2]Department of Computer Science, Memorial University, St. John's, NL, Canada

[3]Division of Community Health and Humanities, Faculty of Medicine, Memorial University, St. John's, NL, Canada

[4]Faculty of Medicine, Memorial University, St. John's, NL, Canada

[5]Faculty of Engineering, Memorial University, St. John's, NL, Canada

[6]Department of Geography, University of Oregon, Eugene, OR, United States

[7]School of Health Administration, Dalhousie University, Halifax, NS, Canada

[*]these authors contributed equally

Corresponding Author:
Daniel Fuller, PhD
School of Human Kinetics and Recreation
Memorial University of Newfoundland
St. John's, NL, A1C 5S7
Canada
Phone: 1 7098647270
Email: dfuller@mun.ca

Abstract

Background: Consumer-wearable activity trackers are small electronic devices that record fitness and health-related measures.

Objective: The purpose of this systematic review was to examine the validity and reliability of commercial wearables in measuring step count, heart rate, and energy expenditure.

Methods: We identified devices to be included in the review. Database searches were conducted in PubMed, Embase, and SPORTDiscus, and only articles published in the English language up to May 2019 were considered. Studies were excluded if they did not identify the device used and if they did not examine the validity or reliability of the device. Studies involving the general population and all special populations were included. We operationalized validity as criterion validity (as compared with other measures) and construct validity (degree to which the device is measuring what it claims). Reliability measures focused on intradevice and interdevice reliability.

Results: We included 158 publications examining nine different commercial wearable device brands. Fitbit was by far the most studied brand. In laboratory-based settings, Fitbit, Apple Watch, and Samsung appeared to measure steps accurately. Heart rate measurement was more variable, with Apple Watch and Garmin being the most accurate and Fitbit tending toward underestimation. For energy expenditure, no brand was accurate. We also examined validity between devices within a specific brand.

Conclusions: Commercial wearable devices are accurate for measuring steps and heart rate in laboratory-based settings, but this varies by the manufacturer and device type. Devices are constantly being upgraded and redesigned to new models, suggesting the need for more current reviews and research.

(JMIR Mhealth Uhealth 2020;8(9):e18694) doi:10.2196/18694

KEYWORDS

commercial wearable devices; systematic review; heart rate; energy expenditure; step count; Fitbit; Apple Watch; Garmin; Polar

Introduction

Globally, physical inactivity is a pressing public health concern. A recent report suggested that about 23% of adults and 81% of school-going adolescents are not meeting physical activity guidelines [1]. Government organizations have attempted to improve these numbers by implementing initiatives aimed at promoting physical activity. Though the successful promotion of physical activity is a complex multifacetted issue, behavior change is a well-established method to increase physical activity [2]. Metrics defining physical activity guidelines from commercial wearable devices have been developed, including 10,000 steps per day [3,4] and 100 steps per minute for moderate to vigorous activity [5]. However, research has shown variation in step count among devices, and the applicability of these metrics may vary by device brand and device type [6].

Research examining consumer wearable devices, such as watches, pendants, armbands, and other accessories, is associated with various labels including Quantified Self [7] and mobile health (mHealth) [8]. These consumer wearable devices are becoming increasingly popular for purchase and use. It has been estimated that in the year 2019, 225 million consumer wearables were sold [9], and studies have suggested that more than a third of adults in Canada and Australia own and use a consumer wearable device [10,11]. Despite their popularity, research is equivocal about whether commercial wearable devices are valid and reliable methods for estimating metrics associated with physical activity including steps, heart rate, and energy expenditure.

In a recent review of 10 articles, Bunn et al [12] noted tendencies of wearables to underestimate energy expenditure, heart rate, and step count. Fitbit wearables were highly correlated with criterion measures of step count during laboratory-based assessment and had consistently high interdevice reliability for both step count and energy expenditure [13]. However, this review found that these devices tended to underestimate energy expenditure, which is consistent with a separate review of Fitbit accuracy [14] indicating that Fitbit wearables provide accurate measures only in limited circumstances.

Commercial wearable devices have the potential to allow for population-level measurement of physical activity and large-scale behavior change. However, questions remain about their reliability and validity. This is especially true of smaller and newer manufacturers of wearable devices for which few or no reliability and validity studies have been conducted. The purpose of this systematic review was to outline and summarize information about the validity and reliability of wearables in measuring step count, heart rate, and energy expenditure in any population. The information summarized herein can be used to inform consumers and can aid researchers in study design when selecting physical activity monitoring devices.

Methods

Design

This systematic review was conducted and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [14]. The review was not registered with PROSPERO. Full-length peer-reviewed original research articles, short reports, and letters to the editor published from January 1, 2000, through May 28, 2019, were included in the search. We limited the search to articles published after the year 2000 because commercial wearable devices were not truly available before that time.

Search Strategy

We conducted a literature search of the following databases: MEDLINE via PubMed (1946 to present); Embase (1947 to present); and SPORTDiscus with full text (1920 to present) via EBSCO. The reference lists of eligible papers were reviewed for additional pertinent references.

A librarian (KR) developed the MEDLINE search strategy, which was peer reviewed by a second librarian according to the Peer Review of Electronic Search Strategies (PRESS) 2015 Guideline Statement [15]. The MEDLINE strategy, which included Medical Subject Heading terms and text words, was translated for the other databases using database-specific controlled vocabulary. We searched the literature using multiple combinations and forms of the following key terms: *accelerometer*, *fitness tracker*, *activity monitor*, *step count*, *wearable device*, *validity*, *reliability*, *accuracy*, *Fitbit*, *Garmin*, *Misfit*, *Jawbone*, *UnderArmour*, *Samsung*, *Apple watch*, *GENEactiv*, *Empatica*, *Mio*, *Amiigo*, *Xiaomi*, *Actigraph*, *Withings*, and *Sensewear* (see Multimedia Appendix 1 for the full search strategies). An English language limit was applied. We included any abstracts and conference proceedings, as well as articles examining any population in the initial search. References were imported into EndNote X8 software (Clarivate Analytics) where duplicate references were removed. The remaining references were then imported into Covidence software (Veritas Health Innovation) for screening.

Study Selection Strategy

The web-based systematic review software Covidence was used for this review. The titles and abstracts of the studies included from the initial database search were independently assessed by at least two authors from the team. Conflicts arising during any step of the screening for inclusion/exclusion were resolved by a third author or by consensus. Following the title and abstract screening, full-text documents of the selected studies were searched and retrieved and were independently assessed for inclusion by at least two authors (EC, JL, and DF). Any conflicts were resolved by discussion and consensus. All reviewers strictly adhered to the defined inclusion criteria.

Eligibility Criteria

Studies that met the following criteria were included in the review: (1) use of any consumer-wearable model from the brand Apple Inc, Empatica, Fitbit, Garmin, Jawbone, Mio, Misfit, Polar, Samsung, UnderArmour, Withings, or Xiaomi; (2) specific examination of the reliability and validity measures of the aforementioned brands; and (3) examination of the device's ability to measure a variable (step count, heart rate, or energy expenditure). Studies with fewer than 10 participants were excluded, as has been done in previous work [13]. Validity of the wearable devices was defined as follows [16]:

- Criterion validity: comparing the devices to a criterion measure of steps, heart rate, or energy expenditure.

Reliability of the trackers included the following [16]:

- Intradevice reliability: consistent test-retest results conducted within the same device.
- Interdevice reliability: consistent results across the same model of wearable device measured at the same time and worn at the same location.

The main exclusion criteria were non-English studies, opinion/magazine articles, and systematic reviews. The initial database search and title/abstract screening included articles examining the accuracy of research-grade wearable devices, but the number of returned results was unmanageable. In order to further elucidate the research question in regard to consumer-wearable devices, before full-text screening, the decision was made to exclude all studies examining the reliability and validity of research-grade devices (Actigraph, GENEactiv, Amiigo, Sensewear Armband, Yamax, Omron, Kenze Lifecorder, Digiwalker, Actical, and Actiheart). Studies in which heart rate and energy expenditure estimates were collected using a chest strap heart rate monitor and transmitted to a wearable device were also excluded. Following text screening, the decision was made to exclude abstracts and conference papers. Following data extraction, the decision was made to exclude all studies examining Jawbone commercial wearables, as the company's application program interface (API) was taken offline in 2018, rendering associated devices defunct. Studies were included in the final review if they had extractable data for the following criterion validity measures: correlation coefficient, group mean or percentage difference, median or mean absolute percentage error (MAPE), or level-of-agreement analysis, or had correlation coefficients for reliability measures. Authors were not contacted if these data were not reported in published or supplementary material. The remaining articles were those that met the inclusion criteria (consumer-grade wearables).

Risk of Bias

In our risk of bias assessment, comparisons that did not report group percentage differences or correlation coefficients (n=192) were excluded from the quantitative analysis. However, rather than exclude these comparisons and studies from the review completely, we included them in a narrative summary of how the measures reported were or were not consistent with exploration of percentage measurement error and correlation.

Data Extraction

We first conducted and documented an in-depth web search of the available consumer-wearable models and their specifications (placement, size, weight, cost, and connectivity). The data extraction process then consisted of the following: (1) categorizing the selected full-text articles into reliability or validity studies (EC, JL, and DF); (2) using a modification of the modified Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) validation subscale used by Feehan et al [13] and an a priori modified COSMIN reliability subscale (Multimedia Appendix 2) to assess the quality and risk of bias of each study (EC and DF); (3) extracting the key characteristics from each selected publication and compiling them into tables. Details from each reviewer were compared, and inconsistencies were resolved through consensus before compiling the results (EC and DF).

Data extracted included characteristics of studies, participants, and devices, including study setting and activity type, outcomes measured, and type of criterion measure used. Correlation coefficients were extracted for all reliability comparisons reported in each study. Correlation coefficients, percentage difference and group mean values, MAPE values, and level-of-agreement data were extracted for all validity comparisons where available. Where group percentage differences were not reported, we calculated group percentage error ($[\text{wearable}_{\text{mean}} - \text{criterion}_{\text{mean}}]/\text{criterion}_{\text{mean}} \times 100$) to allow for comparison across studies. We split a small number of studies (n=10) into "substudies" (n=21), where separate populations were examined in the same publication (see Multimedia Appendix 3 for a more detailed breakdown).

Syntheses

Given the wide range of testing conditions and reported outcomes, we were unable to conduct meta-analyses of the extracted data. We instead conducted a narrative synthesis of the available quantitative data within each examined measure (step count, heart rate, and energy expenditure) using correlation comparisons and group percentage difference as the common metrics for criterion validity and correlation coefficient as the common metric for reliability.

Our interpretation of measurement accuracy was focused on acceptable limits of percentage difference of ±3% in controlled settings and percentage difference of ±10% in free-living settings, as outlined in previous work [13]. We interpreted correlation coefficients as follows: 0 to <0.2, very weak; ≥0.2 to <0.4, weak; ≥0.4 to <0.6, moderate; ≥0.6 to <0.8, strong; and ≥0.8 to 1.0, very strong [17]. We completed all quantitative analyses and plots using RStudio version 1.2.1335 (RStudio Inc) and R version 3.6.0 (The R Foundation).

Secondary analyses explored device brand. Brands were only included in these analyses when the group had 10 or more comparisons available for the measure. Studies that did not report data allowing for the examination of group percentage measurement error were still included in the review if they reported level of agreement or MAPE data. Such studies were included in the risk of bias assessment, the synthesis of study characteristics, and the narrative synthesis of study results.

Availability of Data and Materials

Data are publicly available on the BeapLab Dataverse [18], and the analysis code is available on Github [19].

Results

The initial literature search from the three databases yielded 34,890 unique citations (13,679 [39.21%] from PubMed, 17,560 [50.33%] from Embase, and 3651 [10.46%] from SPORTDiscus). Fourteen additional records were identified through other sources (eg, article reference lists and social media). After duplicate references were removed, 21,083

citations remained. Based on the subsequent title and abstract screening, 20,541 were rejected because they did not meet the inclusion criteria or met the exclusion criteria. Of the 542 that remained for full-text screening, 385 (71.0%) were further excluded for the following reasons: research-grade devices (n=311, 57.4%), wrong variable examined (n=24, 4.4%), fewer than 10 participants (n=14, 2.6%), abstracts (n=13, 2.4%), wrong consumer-grade brand examined (n=10, 1.9%; devices were Yamax, Omron, Kenz Lifecorder, Digiwalker, and uniaxial Actical/Actiheart), no extractable data (n=10, 1.9%), not peer reviewed (n=2, 0.4%), and conference paper (n=1, 0.2%). As a result, a total of 158 publications were included in this systematic review (Figure 1) [14]. Table 1 shows the details of the device brand, model, year, and status (current model or discontinued) in the included studies.

Figure 1. PRISMA flow chart for systematic review of the reliability and validity of commercial wearable devices.

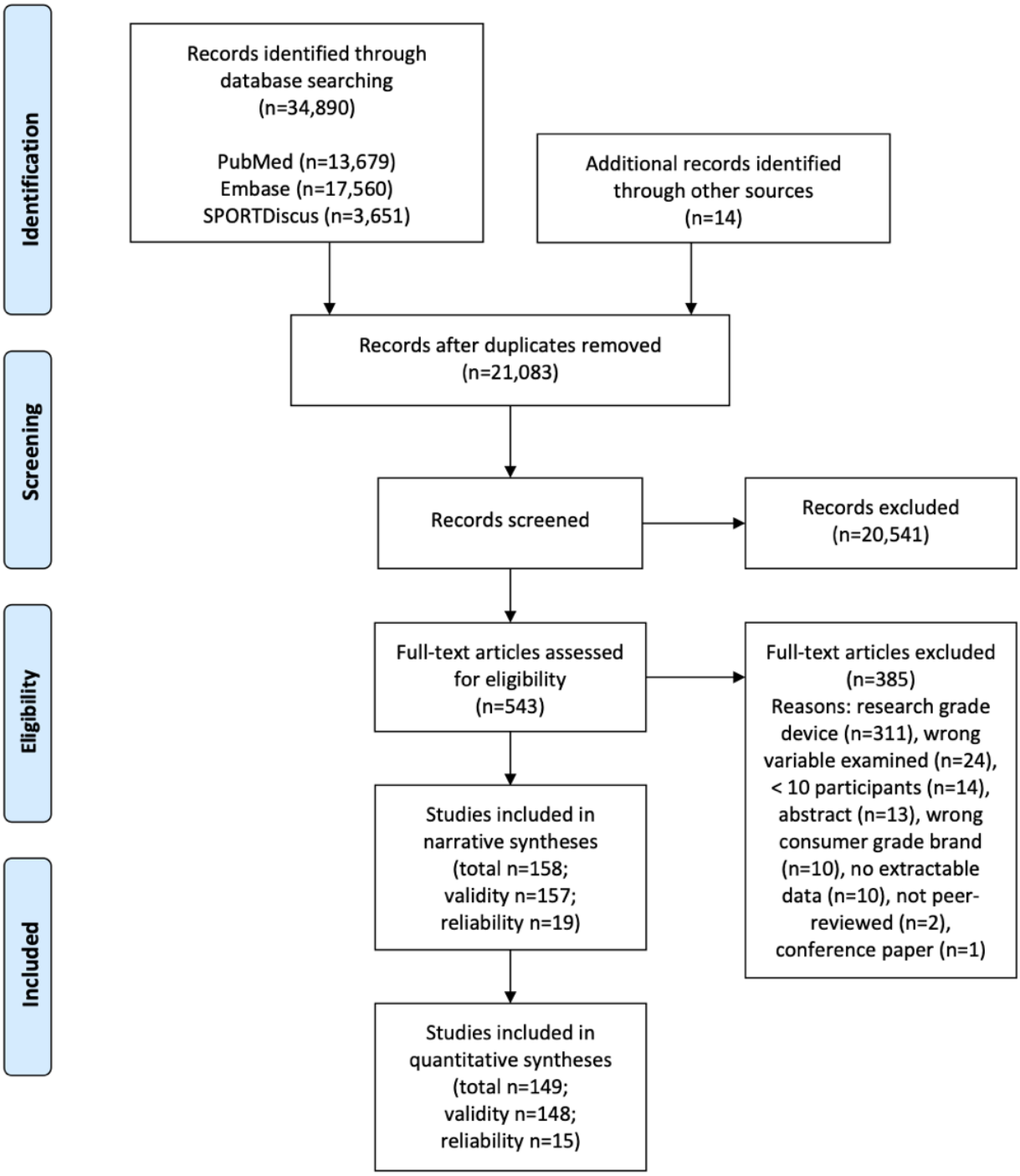


Table 1. Device brand, model, year, current status, wear location, and studies used for the current systematic review.

Brand	Model	Year	Status	Wear location	Studies
Apple	Watch	2015	Discontinued	Wrist	[20-40]
Apple	Watch Series 2	2016	Discontinued	Wrist	[41-44]
Fitbit	Alta	2016	Current model	Wrist	[45]
Fitbit	Blaze	2016	Discontinued	Wrist	[22,40,43]
Fitbit	Charge	2014	Discontinued	Wrist	[45-56]
Fitbit	Charge 2	2016	Discontinued	Wrist	[23,30,43,44,57-63]
Fitbit	Charge HR	2015	Discontinued	Wrist	[20,21,29,32,34,36,38,45,53,64-82]
Fitbit	Classic	2009	Discontinued	Ankle/foot or waist/hip	[83-87]
Fitbit	Flex	2013	Discontinued	Thigh or wrist	[45,50,72,79,80,88-117]
Fitbit	Flex 2	2017	Current model	Wrist	[113]
Fitbit	Force	2013	Discontinued	Wrist	[118,119]
Fitbit	One	2012	Discontinued	Ankle/foot, pant pocket, waist/hip, or wrist	[34,49,52,73,80,88,90,92,93,98,100,102,103,110,116-118,120-138]
Fitbit	Surge	2015	Discontinued	Wrist	[27,35,42,45,54,82,139-143]
Fitbit	Ultra	2011	Discontinued	Chest, pant pocket, upper arm, waist/hip, or wrist	[85,144-148]
Fitbit	Zip	2012	Current model	Ankle/foot, pant pocket, shin, or waist/hip	[34,45,46,51,88,89,92,93,96,103,112,119,127,129,131,141,149-161]
Garmin	Fenix 3 HR	2016	Discontinued	Wrist	[41]
Garmin	Forerunner 225	2015	Discontinued	Wrist	[21,162]
Garmin	Forerunner 235	2015	Current model	Wrist	[40,139,163]
Garmin	Forerunner 405CX	2009	Discontinued	Wrist	[164]
Garmin	Forerunner 735XT	2016	Current model	Wrist	[35]
Garmin	Forerunner 920XT	2014	Discontinued	Wrist	[53]
Garmin	Vivoactive	2015	Discontinued	Wrist	[53]
Garmin	Vivofit	2014	Discontinued	Wrist	[35,46,50,52,53,89,92,104,114,122,130,143,150,159,165-169]
Garmin	Vivofit 2	2015	Discontinued	Wrist	[34,127,170]
Garmin	Vivofit 3	2016	Discontinued	Wrist	[168,171]
Garmin	Vivosmart	2014	Discontinued	Wrist	[32,53,75]
Garmin	Vivosmart HR	2015	Discontinued	Wrist	[36,43,65]
Garmin	Vivosmart HR+	2016	Current model	Wrist	[58,63,140]
Mio	Alpha	2013	Discontinued	Wrist	[25,38,71]
Mio	Fuse	2015	Discontinued	Wrist	[54,64]
Misfit	Flash	2015	Discontinued	Waist/hip	[32]
Misfit	Shine	2012	Discontinued	Ankle/foot, chest, pant pocket, waist/hip, or wrist	[74,79,89,96,99,104,105,131,159,169]

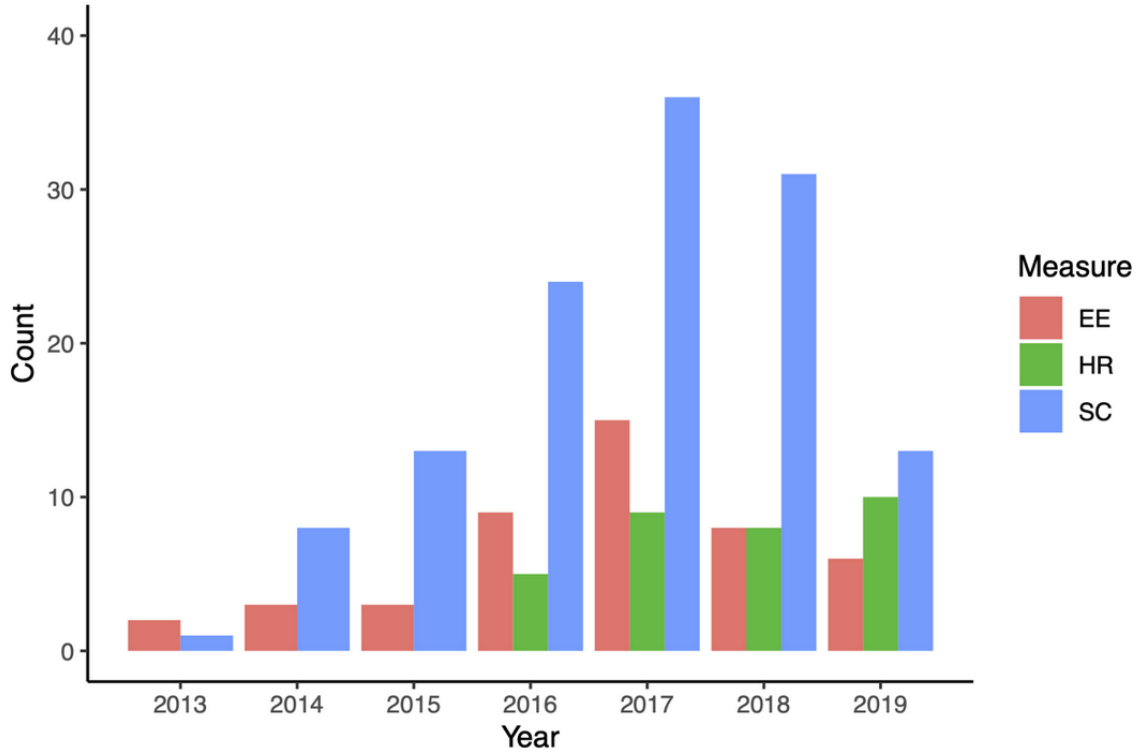
Brand	Model	Year	Status	Wear location	Studies
Polar	A300	2015	Discontinued	Wrist	[172]
Polar	A360	2015	Discontinued	Wrist	[25,43,140]
Polar	Active	2011	Discontinued	Wrist	[173]
Polar	Loop	2013	Discontinued	Wrist	[32,50,53,79,89,167]
Polar	M600	2016	Current	Wrist	[56]
Polar	V800	2016	Discontinued	Wrist	[174]
Samsung	Gear 2	2014	Discontinued	Wrist	[140]
Samsung	Gear S	2014	Discontinued	Wrist	[32,38]
Samsung	Gear S2	2015	Discontinued	Wrist	[35]
Samsung	Gear S3	2016	Discontinued	Wrist	[42,44]
Withings	Pulse	2013	Discontinued	Collar, pant pocket, waist/hip, or wrist	[89,96,131,166]
Withings	Pulse O2	2013	Discontinued	Collar, waist/hip, or wrist	[104,122,123,169,175]
Withings	Pulse Ox	2014	Current model	Waist/hip or wrist	[53,58]
Xiaomi	Mi Band	2014	Discontinued	Wrist	[42]
Xiaomi	Mi Band 2	2016	Discontinued	Wrist	[81]

Study and Participant Characteristics

Of the 158 publications included, 143 were full-text research articles, 10 were brief reports, and five were letters to the editor. Publication year ranged from 2013 to 2019, with the amount of publications increasing from 2013 to 2017 (2013, n=2; 2014, n=8; 2015, n=11; 2016, n=30; 2017, n=43). We also included an additional 40 and 24 studies published in 2018 and 2019, respectively.

Within those 158 publications, 169 studies/substudies were identified. Among these, 168 (99.4%) examined validity and 19 (11.2%) examined reliability. Moreover, 126 studies examined step count (125 validity and 16 reliability), 32 examined heart rate (32 validity and 3 reliability), and 43 examined energy expenditure (42 validity and 5 reliability) (Figure 2). Furthermore, 130 examined populations in a controlled environment and 48 examined populations in a free-living environment. A total of 1838 comparisons were identified, of which 166 examined reliability (mean 8, SD 11 per reliability study; range 1-40) and 1672 examined validity (mean 10, SD 15 per validity study; range 1-98).

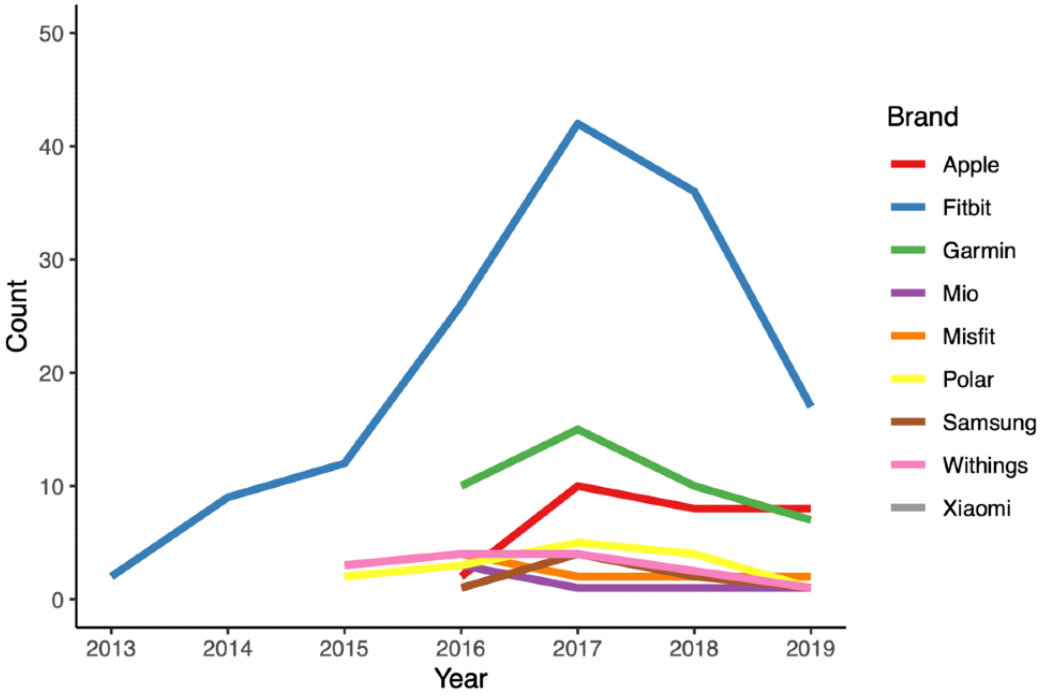
Figure 2. Number of studies published per year by measurement type. EE: energy expenditure; HR: heart rate; SC: step count.



The 169 studies/substudies comprised a total of 5934 participants, with a mean of 35 (SD 27) participants per study (range 10-185). One hundred and sixty-one studies reported sex, and 51.08% (2861/5601) of participants were female. One hundred and fifty-eight studies reported age, with a mean participant age of 36.8 years (SD 18.3; range 3.7-87 years). One hundred and fifty-nine studies examined adult populations (age ≥18 years) and 10 studies examined children. One hundred and thirty-three studies included only healthy participants, while the other 36 studies included participants with mobility limitations and/or chronic diseases (Multimedia Appendix 4).

Fitbit consumer-grade wearables were examined most frequently (144 studies examining 12 models), followed by Garmin (42 studies, 13 models), Apple (28 studies, 2 models), Polar (15 studies, 6 models), Misfit (13 studies, 2 models), Withings (12 studies, 2 models), Samsung (8 studies, 4 models), Mio (6 studies, 2 models), and Xiaomi (2 studies, 2 models) (a complete list of examined models is provided in Multimedia Appendix 5) (Figure 3). Wearables were typically examined while worn on the wrist (n=131, examining at least one wrist-worn device) or at the waist/hip (n=71, locations included the waist, hip, belt, and pants pocket). Substantially fewer studies examined wearables worn on the torso (n=14, locations included the chest, bra, lanyard, and shirt collar) and lower limb (n=13, locations included the thigh, shin, ankle, and foot).

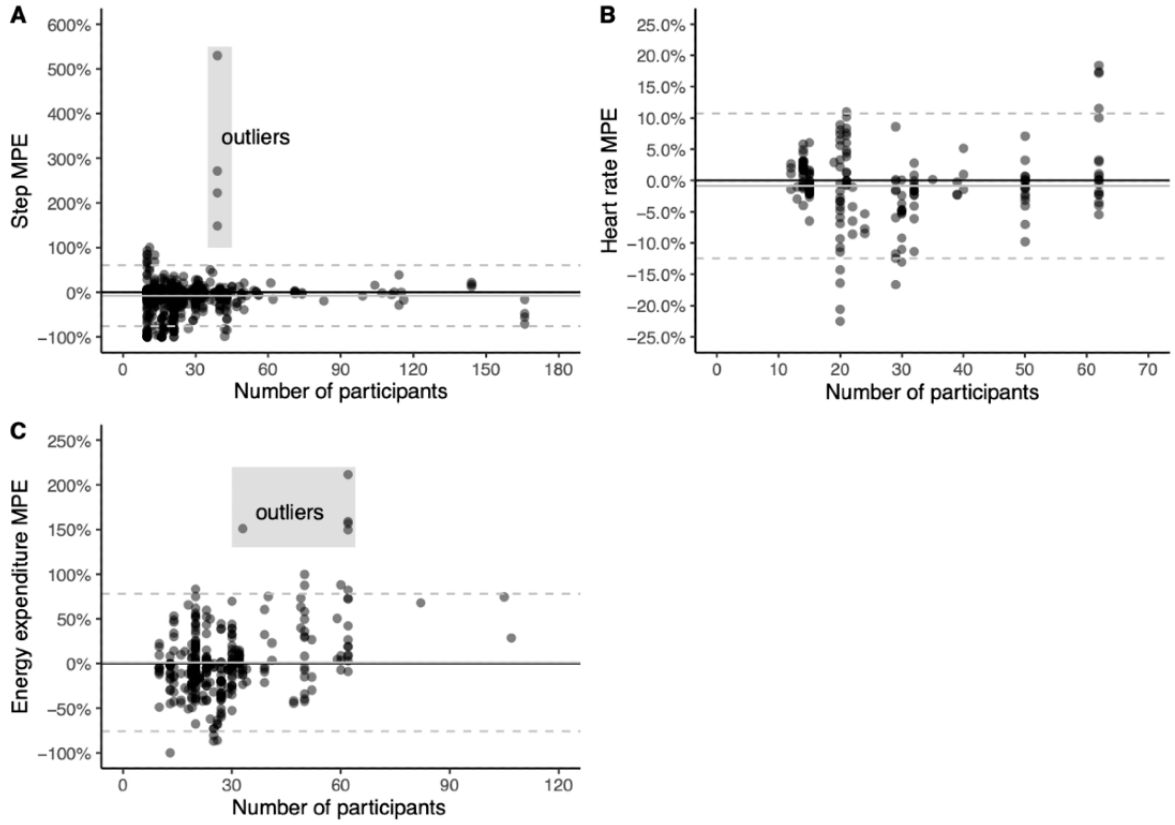
Figure 3. Line graph of studies published per year by device brand.



Risk of Bias

Of 169 studies, 140 (82.8%; 1640 of 1838 [89.23%] comparisons) were rated fair or poor for sample size (<50 participants), but were not excluded from the analysis owing to the paucity of studies with excellent (≥100 participants, n=7) and good (50-99 participants, n=22) sample sizes. We additionally explored the potential for bias related to sample size in step count, heart rate, and energy expenditure by examining the percentage error dispersion by sample size using scatter plots (Figure 4).

Figure 4. Mean percentage error (MPE) plots by study sample size for step count, heart rate, and energy expenditure. The solid black line represents zero. The solid grey line represents average MPE for all data points. The dashed grey lines represent the 95% CIs.



In these examinations, we saw no apparent systematic bias for measurement error beyond a small number of comparisons showing extreme overestimation (four comparisons in step count and five comparisons in energy expenditure). The four extreme outliers for step count involved measurement during sedentary and light physical activity in a single study with fewer than 40 participants [20] and were likely inflated by the limited number of steps accumulated during those bouts. As a result, we excluded these four comparisons from the quantitative syntheses. Upon closer examination of the five extreme outliers for energy expenditure (four occurred in a study with greater than 60 participants [21] and one occurred in a study with fewer than 40 participants [41]), we determined that these were likely true reflections of tendencies to overestimate energy expenditure during sedentary and low-intensity activities, and therefore, we included these five comparisons in the quantitative syntheses.

Validity: Controlled Settings

We examined criterion validity for step count, heart rate, and energy expenditure separately for controlled and free-living settings. For controlled settings, we also had sufficient data to examine validity by brand and devices within brands.

Validity for Step Count in Controlled Settings

A total of 90 studies (979 comparisons) examined wearable device step count measurements compared with reference standard criterion measures of manual counting [32,34-38,42,46,47,50-53,57,58,72,80-84,88-102,109,114-125,138-141,144-147,149-153,158-161,165,169-171,173] and accelerometry [20,60,64-66,85,103,109,126-128,148,154,164] (Multimedia Appendix 6). Of these, 67 studies recruited healthy adults (mean age 35.4 years, SD 17.4 years), 20 studies recruited adults living with limited mobility/chronic diseases (mean age 60.1 years, SD 10.5 years), two studies recruited children living with limited mobility/chronic diseases (mean age 12.5 years, SD 2.9 years), and one study recruited healthy children (mean age 3.7 years, SD 0.6 years). Wearable devices were worn on the lower limb (foot, ankle, shin, and thigh), torso, waist/hip, and wrist.

Group measurement error was reported or calculable for 805 of the 979 comparisons, regardless of the criterion measure. Of these, 45.2% (n=364) were within ±3% measurement error, 42.7% (n=344) were below −3% measurement error, and 12.1% (n=97) were above 3% measurement error. The overall tendency was to underestimate step count (mean: −9%, median: −2%).

Validity for Heart Rate in Controlled Settings

A total of 29 studies (266 comparisons) examined wearable device heart rate measurements compared with reference standard criterion measures, including electrocardiography [22,23,38-40,43,44,54,61,62,67-70,142,162,176], Polar brand chest straps [20,21,24-28,58,63,71,163], and pulse oximetry [66], in controlled settings (a detailed list of the criterion measures used is presented in Multimedia Appendix 6). Of these, 24 studies recruited healthy adults (mean age 29.8 years, SD 10.5 years), four studies recruited adults living with limited mobility/chronic diseases (mean age 59.6 years, SD 9.0 years), and one study recruited children undergoing surgery (mean age 8.2 years, SD 3.1 years). All wearable devices were worn on the wrist.

Group measurement error was reported or calculable for 177 of 266 comparisons, regardless of the criterion measure. Of these, 56.5% (n=100) were within ±3% measurement error, 24.9% (n=44) were below −3% measurement error, and 18.6% (n=33) were above 3% measurement error. There was a slight overall tendency toward underestimation of heart rate (estimated median error: −1%).

Validity for Energy Expenditure in Controlled Settings

A total of 36 studies (312 comparisons) examined wearable device energy expenditure measurements compared with reference standard criterion measures, including direct calorimetry [86,104] and indirect calorimetry [20,21,29-31,38,39,41-43,53,55,63,66,73,85,87,93,95,97,103,105,116,117,129,130,142,143,146,148,159,165,166,177], in controlled settings. Of these, 35 studies recruited healthy adults (mean age 27.2 years, SD 7.1 years), and one study recruited adults living with cardiovascular disease (mean age 64.2 years, SD 2.3 years). Wearable devices were worn on the wrist, waist/hip, and torso.

Group measurement error was reported or calculable for 305 of the 312 comparisons, regardless of the criterion measure. Of these, 9.2% (n=28) were within ±3% measurement error, 54.1% (n=165) were below −3% measurement error, and 36.7% (n=112) were above 3% measurement error. Studies showed a tendency to underestimate energy expenditure and to provide inaccurate measures of energy expenditure compared with the criterion.

Validity in Controlled Settings by Brand

Figure 5 shows the mean percentage error (MPE) for step count, heart rate, and energy expenditure by device brand for devices with 10 or more comparisons. Figure 6 shows the MPE for step count, heart rate, and energy expenditure by device brand and model for devices with 10 or more comparisons.

Figure 5. Box plots representing mean percentage error (MPE) for steps, heart rate, and energy expenditure by device brand for devices with 10 or more comparisons.

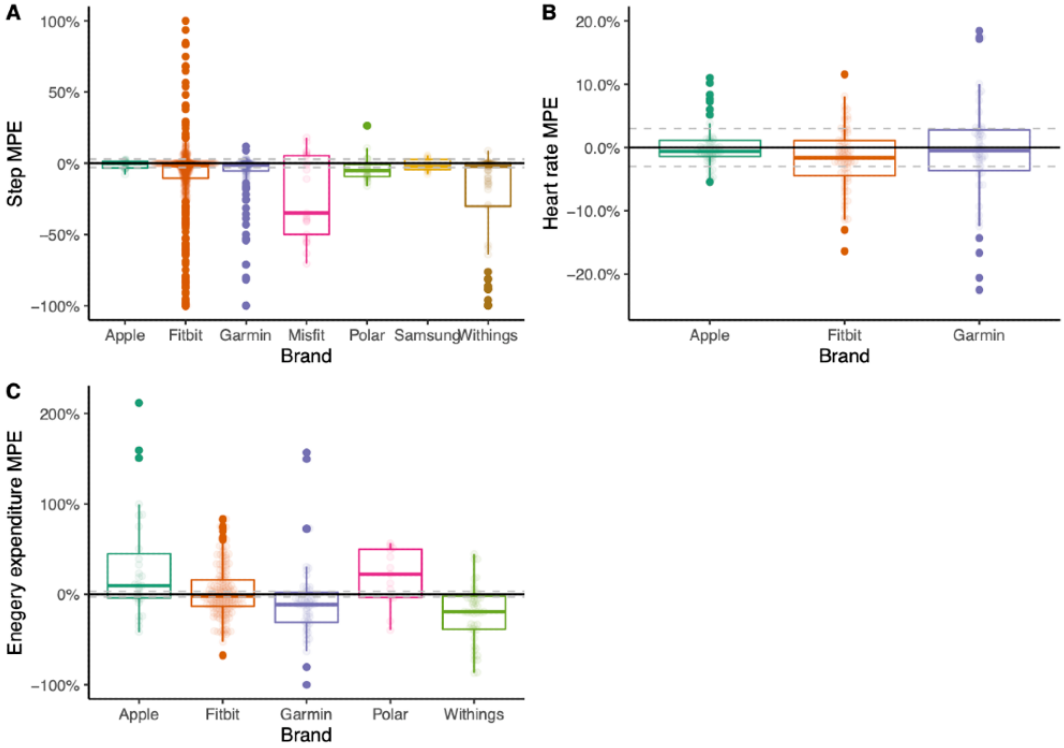
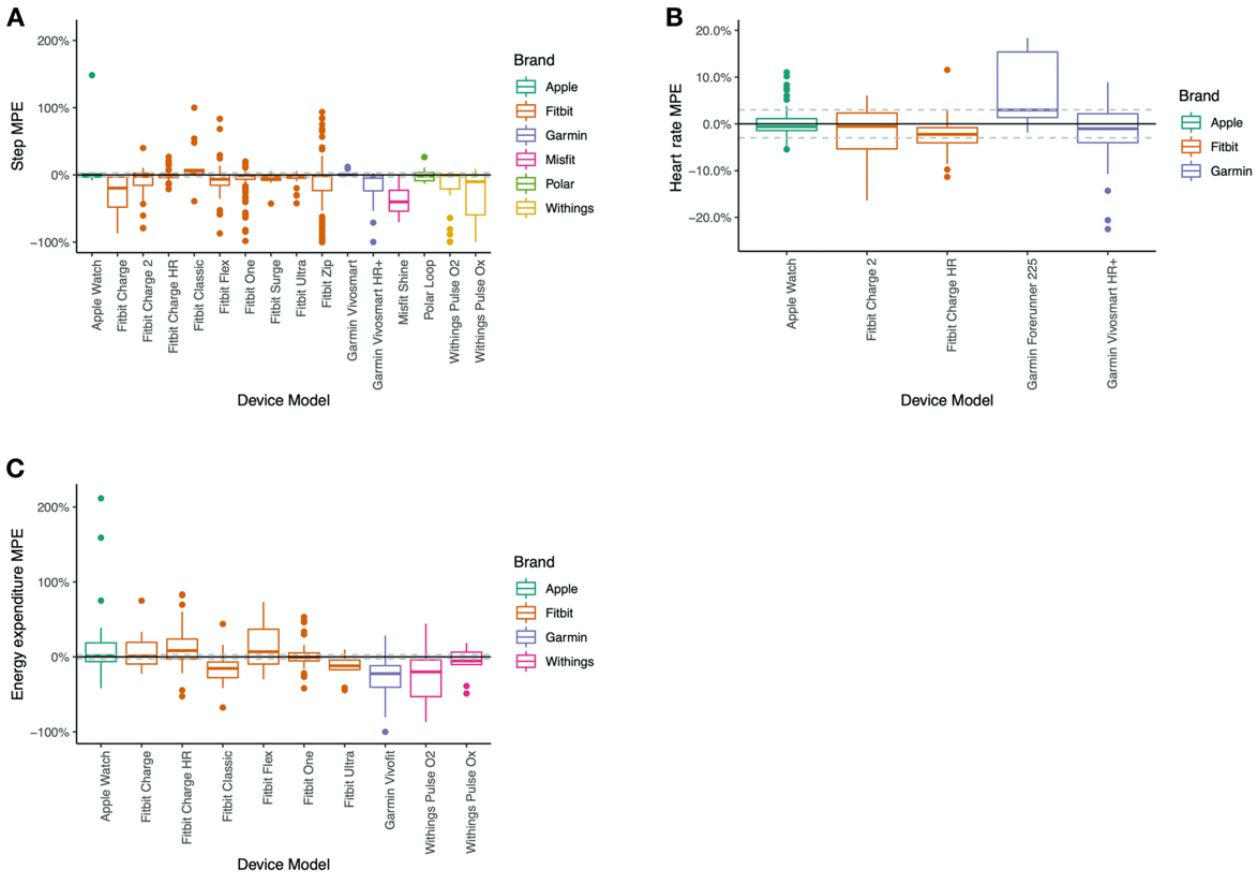


Figure 6. Box plots representing mean percentage error (MPE) for steps, heart rate, and energy expenditure by device brand and model for devices with 10 or more comparisons.



Validity for Step Count by Brand

We observed that the error level varied by device brand (Figure 5). Withings and Misfit wearables consistently underestimated step count, and Apple and Samsung had less measurement variability than other brands. There are possible interactions between the number and size of studies and device wear location that may influence the brand comparisons. For example, Apple Watch and Samsung have the tightest ranges for step count estimates but have relatively fewer studies compared with other brands.

Validity for Heart Rate by Brand

For heart rate, measurement error also varied by device brand (Figure 5). Apple Watch was within ±3% 71% (35/49) of the time, while Fitbit wearables were within ±3% 51% (36/71) of the time and Garmin wearables were within ±3% 49% (23/47) of the time. Despite similar ±3% measurement error rates, Fitbit appeared to underestimate heart rate more than Apple Watch and Garmin.

Validity for Energy Expenditure by Brand

For energy expenditure estimates, no brand of wearable was within ±3% measurement error more than 13% of the time (Figure 5). Underestimation of energy expenditure (less than −3%) was observed in Garmin wearables 69% (37/51) of the time and in Withings wearables 74% (34/46) of the time. Conversely, Apple wearables overestimated energy expenditure 58% (18/31) of the time and Polar wearables overestimated energy expenditure 69% (9/13) of the time. Fitbit devices tended to provide inaccurate measures compared with the criterion, underestimating 48.4% (76/157) of the time and overestimating 39.5% (62/157) of the time, despite the boxplot in Figure 5 showing a reasonable median value for accuracy.

Validity: Free-Living Settings

There were relatively few studies on wearable device validity in free-living conditions. Fitbit was the only brand with more than 10 studies published for step count validity in free-living conditions, and no brands had more than 10 studies for heart rate or energy expenditure. As a result, we have not shown plots of MPE for free-living conditions.

Validity for Step Count in Free-Living Settings

A total of 42 studies (84 comparisons) examined wearable device step count measurements compared with the reference standard criterion measure of accelerometry [33,45,48,49,56,59,60,64,74-76,89,96,101,106-112,120,131-136,149,154-156,159,167,168,172-175] in free-living settings (Multimedia Appendix 6). Of these, 28 studies recruited healthy adults (mean age 33.7 years, SD 13.9 years), nine studies recruited adults living with limited mobility/chronic diseases (mean age 60.1 years, SD 11.2 years), four studies recruited healthy children (mean age 12.5 years, SD 2.6 years), and one study recruited children living with cardiac diseases (mean age 13 years, SD 2.2 years). Wearable devices were worn on the lower limb (foot, ankle, and shin), torso, waist/hip, and wrist.

Group measurement error was reported or calculable for 69 of the 84 comparisons, regardless of the criterion measure. Of these, 42% (n=29) were within ±10% measurement error, 17% (n=12) were below −10% measurement error, and 41% (n=28) were above 10% measurement error. The overall tendency was slight overestimation of step count (mean: 5%, median: 6%). Among the remaining comparisons, 11 of 15 reported MAPE, of which 40% (n=6) were below 10% measurement error and 60% (n=9) were above 10% measurement error.

Validity for Heart Rate in Free-Living Settings

Three studies (five comparisons) examined wearable device heart rate compared with the reference standard criterion measure of a Polar brand chest strap in free-living settings [75,77,78]. Of these, one study recruited healthy adults (mean age 25.4 years, SD 3.7 years), one study recruited healthy children (mean age 8 years, SD 1.8 years), and one study recruited adults recovering from stroke (mean age 64.4 years, SD 15 years). All wearable devices were worn on the wrist. Group measurement error was reported or calculable for one of the five comparisons, with the Fitbit Charge HR falling within ±10% measurement error in the study examining healthy children. Three of the four remaining comparisons examined the Fitbit Charge HR in adults and noted underestimation of heart rate that varied depending on activity intensity, but all reported that MAPE values fell within 10% measurement error. Correlation coefficients were strong to very strong in four of the five comparisons and moderate in one comparison examining estimation during high-intensity activity.

Validity for Energy Expenditure in Free-Living Settings

Nine studies (22 comparisons) examined energy expenditure in free-living settings compared with the criterion measures of doubly labeled water [104] and accelerometry [29,49,79,101,131,172,174,175]. Eight studies recruited healthy adults (mean age 27.7 years, SD 3.8 years) and one study recruited adults with chronic obstructive pulmonary disease (mean age 66.4 years, SD 7.4 years). Wearable devices were worn on the wrist or waist/hip.

Group measurement error was reported or calculable for 17 of the 22 comparisons, regardless of the criterion measure. Of these, 18% (n=3) were within ±10% measurement error, 53% (n=9) were below −10% measurement error, and 29% (n=5) were above 10% measurement error. There was an overall tendency to underestimate energy expenditure (mean: −3%, median: −11%). Xiaomi data were not analyzed in a single indirect calorimetry study owing to the lack of data [53].

Reliability

Nineteen studies (166 comparisons) with sample sizes ranging from 11 [94] to 56 [151] reported inter- or intradevice reliability for Apple (seven comparisons), Fitbit (92 comparisons), Garmin (22 comparisons), Polar (one comparison), and Withings (44 comparisons). The majority of comparisons (153/166) reported interdevice reliability for step count, heart rate, or energy expenditure. No studies reported intradevice reliability for heart rate or energy expenditure. We have not reported between-brand comparisons for inter- or intradevice reliability owing to the small number of comparisons for each brand.

Interdevice Reliability for Step Count

Twelve studies (51 comparisons) with sample sizes ranging from 13 [117,138] to 56 [151] reported on interdevice reliability for step count [50,58,72,85,94,110,113,116,117,121,125,138,151,161,171]. The majority of correlation coefficients for step count interdevice reliability were very strong (n=35), with only a small number (n=3) being reported as strong.

Intradevice Reliability for Step Count

Two studies (13 comparisons) reported on intradevice reliability for step count, with sample sizes of 20 [82] and 24 [150]. Intradevice reliability correlations were very weak (n=1), weak (n=2), moderate (n=5), strong (n=2), and very strong (n=3). The mean correlation coefficient was 0.58.

Interdevice Reliability for Heart Rate

Three studies (23 comparisons) examined interdevice reliability for heart rate [24,26,58], with analyzed sample sizes ranging from 13 [24] to 21 [26]. Apple Watch showed very good interdevice reliability at 5-s epochs during treadmill bouts at 4, 7, and 10 km/h, with reliability increasing and standard typical error decreasing with increasing pace [26]. Similar standard typical error levels were seen in maximum heart rate measured during a single incremental maximal oxygen uptake test performed on a treadmill and heart rate taken from the highest 30-s mean heart rate, with somewhat lower correlation coefficients [24]. In the examination of interdevice reliability in healthy older adults, Fitbit Charge 2 showed good reliability during treadmill and overground bouts and poor reliability during hand movement tasks such as dusting [58]. During the same tasks, Garmin Vivosmart HR+ showed good reliability during all tasks and had narrower limits of agreement than Fitbit.

Interdevice Reliability for Energy Expenditure

Five studies (50 comparisons) reported on interdevice reliability [85,113,116,117,166], with analyzed sample sizes ranging from 13 [117] to 29 [113]. All five studies recruited healthy adults (mean age 26.3 years, SD 3.9 years). Correlation coefficients were reported for 16 of 50 comparisons. Of these, 13% (n=2) were rated very weak, 6% (n=1) were rated moderate, 6% (n=1) were rated strong, and 75% (n=12) were rated very strong.

Discussion

Overview

The purpose of this study was to examine the validity and inter- and intradevice reliability of commercial wearable devices in measuring steps, heart rate, and energy expenditure. Our review focused on both a breadth of devices and reproducibility. Our review included nine brands and 45 devices with the number of comparisons ranging from 201 for the Fitbit Zip to one for the Garmin Forerunner 405CX and the Polar M600. For comparison, two recent reviews from 2017 included two brands and 16 devices [13] and seven brands and eight devices [79]. A review from 2016 included eight devices [32]. Along with this review, we have published our dataset and code to reproduce our findings.

Our bias assessment showed no apparent bias toward studies of different sample sizes. However, there is a strong overrepresentation of studies with 20 participants. There were some outliers in our findings; however, considering the number of included comparisons, this is to be expected.

Reliability and Validity

Criterion validity of commercial wearables varied by study type (controlled or free-living), brand, and device. For step count, our review showed that in controlled laboratory settings, a higher proportion of devices showed accuracy, and this was within a tighter limit of acceptable accuracy compared with free-living conditions. In both controlled and free-living studies, when not correctly estimating steps, devices tended to underestimate values. Validity compared with criteria was the best for Apple Watch and Garmin, while the MPE values for Fitbit, Samsung, and Withings fell within ±3% on average. Within brands, devices appeared to vary, with Fitbit Classic tending to overestimate steps, while Fitbit Charge tending to underestimate steps; however, the variability observed could be attributed to differences in the number of comparisons for each device and in wear locations of the devices. Our findings are consistent with previous reviews [178].

In controlled settings across all devices, heart rate was accurately measured with only a very small tendency for underestimation. Heart rate validity was only sufficiently tested in Apple Watch, Fitbit, and Garmin devices. Heart rate measured by photoplethysmography is only available in relatively new commercial wearable devices. All of the brands measured heart rate to within ±3% on average in controlled settings. There were few studies examining the validity of heart rate measures in free-living conditions, but it appears that Fitbit devices may underestimate heart rate depending on activity intensity. All devices were within acceptable measurement error for heart rate. To our knowledge, this is the first systematic review to examine heart rate validity, and it appears that devices are able to measure heart rate within acceptable limits.

Energy expenditure estimates varied widely with less than 10% of estimates falling within acceptable limits in controlled settings. In many of the studies, there did appear to be a tendency for systematic over or underestimation. On average, only Fitbit measured energy expenditure to within acceptable limits, but there was wide variation around the estimate. Energy expenditure estimates also varied by model, with the Fitbit Classic underestimating the value considerably and Fitbit Charge HR overestimating the value. We hypothesize that Fitbit may provide the best, though still not acceptable, measure of energy expenditure because the algorithm employs a published equation for estimating resting metabolic rate [179]. To our knowledge, the other brands do not publish information about the energy expenditure estimates. There does not appear to be a relationship among more accurate estimates of energy expenditure in devices that include heart rate (Multimedia Appendix 7).

Interdevice reliabilities for steps, heart rate, and energy expenditure were all very strong. However, compared with validity studies, there were fewer reliability studies, and we were not able to conduct comparisons between brands or devices owing to small sample sizes. Sufficient data for intradevice

reliability was only available for step count. The results showed considerable variability within the same device for step count for Fitbit Charge HR, Fitbit Surge, Fitbit Zip, and Garmin Vivofit, with five, five, one, and two comparisons, respectively.

Future Research

Future research in this area should focus on the following three main topics: relevance and age of the devices tested, data acquisition from the devices, and algorithms used by companies. First, relevance of the devices is important. Owing to rapidly developing technology, the majority of the tested devices included in this review are now out of date or discontinued. The nature of the consumer technology market is such that updated product iterations are commissioned even before the original iteration of a device is released. For example, the newest Apple Watch included in the review is the Series 2 watch. The Series 5 watch was released in the fall of 2019. The results are similar for all devices and brands; the Fitbit Charge HR is a popular model for validity and reliability studies, likely because of its moderate price point (approximately US $150) compared with more expensive models (eg, Garmin Fenix 5, approximately US $500). Given the current device specialization, device relevance (eg, swimming or sleep-specific watches), and price difference between devices, continuing to conduct the types of reliability and validity studies reported here will be a challenge. The increasing pace of device release combined with device specialization makes this type of research challenging.

Second, few studies reported on how data were acquired from the devices. We believe this has implications for the scale of and usability of the data collected. For example, in order to collect data, we infer that some studies counted the steps recorded on the device in short time intervals instead of connecting the device to a platform after recording. Other studies exported and downloaded data from user accounts on the brand website, while others collected data from the brand API. Collecting data from the device API is the best and most scalable method for physical activity researchers when using wearable device data. In order to do so, we must develop interdisciplinary collaborations and open source tools to allow these data to be collected (eg, Open mHealth) [180].

Third, the algorithms used in consumer wearables are constantly changing based on sensor development and technological advances. Companies can update their devices' firmware and algorithm at any time. When the device is synced, the firmware is updated. Feehan et al discussed the importance of firmware updates in their review [13]. While we believe this is important, it is clear that companies must be more open about the algorithms they are using to estimate steps, heart rate, and energy expenditure. Given the continuing release of new devices, firmware and algorithm updates to existing devices, and lack of availability of raw data, we believe researchers may need to shift focus from traditional reliability and validity research to studies that can provide open estimates for physical activity intensities or sleep standardized across devices. These studies will need to use device APIs and machine learning methods in collaboration with interdisciplinary teams in order to move the field forward.

Limitations

Over the course of time that it took to complete this review, much has changed with market share, technology, and even research methodologies. Though the market share of companies was a large determining factor of what devices were included in this review, the consumer wearable market is volatile. On November 1, 2019, Google purchased Fitbit for US $2.3 billion, a massive shift for the consumer wearable device market [181,182]. Further to this limitation is the ever-changing nature of consumer technology. As Table 1 shows, many of the devices utilized in the studies included in this review are so out of date that they are no longer available on the market. There is some potential for bias when including only English language studies in systematic reviews. However, studies have shown that the effect may be small in general but may be difficult to measure for an individual systematic review [183,184].

Conclusion

This systematic review of 158 publications included assessments of consumer wearable devices from nine brands (Apple Inc, Fitbit, Garmin, Mio, Misfit, Polar, Samsung, Withings, and Xiaomi), with a focus on the reliability and validity of the devices in measuring heart rate, energy expenditure, and step count. This review examined the validity of consumer wearable devices in free-living and laboratory settings and further highlighted results of the inter- and intradevice reliability of the nine consumer wearable brands. Among the studies included, Fitbit was studied the most and Xiaomi and Mio were studied the least. Apple and Samsung had the highest validity for step count, and Apple, Fitbit, and Garmin were accurate nearly 50% of the time. No brand fell within the acceptable accuracy limits for energy expenditure. Interdevice reliabilities for steps, heart rate, and calories were all very strong. Sufficient data for intradevice reliability were only available for step count, and the results showed considerable variability. There was no specific device or brand that involved a complete assessment across all measures, and no specific brand stood out as the "gold standard" in fitness wearables. This review highlights the validity and reliability of readily available wearable devices from brands and serves to guide researchers in making decisions about including them in their research. As new devices and models enter the market, up-to-date documentation can help direct their use in the research setting.

Acknowledgments

The authors would like to thank Kristen Romme, a librarian at Memorial University, for writing and executing the searches. They would also like to thank Dr Sherri Tran who assisted with article screening as part of her medical training. Funding for this research was provided by Dr Fuller's Canada Research Chair (#950-230773).

Authors' Contributions

DF, EC, and JF conceptualized the paper and lead the review. EC, JF, KO, MT, BS, RB, DvH, LS, HL, and NT conducted screening. DF, EC, JF, and KC conducted methodological review and data analysis. All authors contributed to writing of the manuscript and approved the submitted version of the manuscript.

Conflicts of Interest

Author JL was employed by Garmin Inc. during the publication process but after completion of the paper. All other authors have no conflicts to declare.

Multimedia Appendix 1
Search strategies.
[PDF File (Adobe PDF File), 35 KB - mhealth_v8i9e18694_app1.pdf]

Multimedia Appendix 2
Modified Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) risk of bias.
[PDF File (Adobe PDF File), 131 KB - mhealth_v8i9e18694_app2.pdf]

Multimedia Appendix 3
Substudies.
[PDF File (Adobe PDF File), 11 KB - mhealth_v8i9e18694_app3.pdf]

Multimedia Appendix 4
Study demographics.
[PDF File (Adobe PDF File), 254 KB - mhealth_v8i9e18694_app4.pdf]

Multimedia Appendix 5
Device models.
[PDF File (Adobe PDF File), 197 KB - mhealth_v8i9e18694_app5.pdf]

Multimedia Appendix 6
Criterion measures.
[PDF File (Adobe PDF File), 193 KB - mhealth_v8i9e18694_app6.pdf]

Multimedia Appendix 7
Energy expenditure mean absolute percentage error stratified by heart rate measurement on the device.
[PDF File (Adobe PDF File), 5 KB - mhealth_v8i9e18694_app7.pdf]

References

1. World Health Organization. Global recommendations on physical activity for health. World Health Organization. URL: https://www.who.int/dietphysicalactivity/factsheet_recommendations/en/ [accessed 2020-08-17]
2. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. Am J Public Health 2011 Apr;101(4):751-758 [FREE Full text] [doi: 10.2105/AJPH.2010.194381] [Medline: 21330590]
3. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De CK, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
4. Schneider PL, Bassett DR, Thompson DL, Pronk NP, Bielak KM. Effects of a 10,000 steps per day goal in overweight adults. Am J Health Promot 2006;21(2):85-89. [doi: 10.4278/0890-1171-21.2.85] [Medline: 17152246]
5. Marshall SJ, Levy SS, Tudor-Locke CE, Kolkhorst FW, Wooten KM, Ji M, et al. Translating physical activity recommendations into a pedometer-based step goal: 3000 steps in 30 minutes. Am J Prev Med 2009 May;36(5):410-415. [doi: 10.1016/j.amepre.2009.01.021] [Medline: 19362695]
6. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
7. Almalki M, Gray K, Sanchez FM. The use of self-quantification systems for personal health information: big data management activities and prospects. Health Inf Sci Syst 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S1 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S1] [Medline: 26019809]

8. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. Critical Public Health 2013 Dec;23(4):393-403. [doi: 10.1080/09581596.2013.794931]
9. Costello K. Gartner Says Worldwide Wearable Device Sales to Grow 26 Percent in 2019. Gartner, Inc. 2018. URL: https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow- [accessed 2020-08-17]
10. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open 2016 Jul 07;6(7):e011243 [FREE Full text] [doi: 10.1136/bmjopen-2016-011243] [Medline: 27388359]
11. Macridis S, Johnston N, Johnson S, Vallance JK. Consumer physical activity tracking device ownership and use among a population-based sample of adults. PLoS One 2018;13(1):e0189298 [FREE Full text] [doi: 10.1371/journal.pone.0189298] [Medline: 29293532]
12. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. Int J Exerc Sci 2018;11(7):503-515 [FREE Full text] [Medline: 29541338]
13. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
14. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
15. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
16. Higgins PA, Straub AJ. Understanding the error of our ways: mapping the concepts of validity and reliability. Nurs Outlook 2006;54(1):23-29. [doi: 10.1016/j.outlook.2004.12.004] [Medline: 16487776]
17. Swinscow T, Campbell M. Statistics at Square One. Oxford: BMJ Books; 2009.
18. Replication Data for: Systematic Review of the Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate. Harvard Dataverse. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/O7GQIM [accessed 2020-08-06]
19. wearable_systematic_review. GitHub. URL: https://github.com/walkabillylab/wearable_systematic_review [accessed 2020-08-06]
20. Bai Y, Hibbing P, Mantis C, Welk G. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018 Aug;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
21. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
22. Koshy AN, Sajeev JK, Nerlekar N, Brown AJ, Rajakariar K, Zureik M, et al. Smart watches for heart rate assessment in atrial arrhythmias. Int J Cardiol 2018 Sep 01;266:124-127. [doi: 10.1016/j.ijcard.2018.02.073] [Medline: 29887428]
23. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 2019 Jun;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]
24. Abt G, Bray J, Benson AC. The validity and inter-device variability of the Apple Watch™ for measuring maximal heart rate. J Sports Sci 2018 Jul;36(13):1447-1452. [doi: 10.1080/02640414.2017.1397282] [Medline: 29090987]
25. Bunn J, Wells E, Manor J, Webster M. Evaluation of Earbud and Wristwatch Heart Rate Monitors during Aerobic and Resistance Training. Int J Exerc Sci 2019;12(4):374-384 [FREE Full text] [Medline: 30899350]
26. Khushhal A, Nichols S, Evans W, Gleadall-Siddall D, Page R, O'Doherty A, et al. Validity and Reliability of the Apple Watch for Measuring Heart Rate During Exercise. Sports Med Int Open 2017 Oct;1(6):E206-E211 [FREE Full text] [doi: 10.1055/s-0043-120195] [Medline: 30539109]
27. Pope ZC, Lee JE, Zeng N, Gao Z. Validation of Four Smartwatches in Energy Expenditure and Heart Rate Assessment During Exergaming. Games Health J 2019 Jun;8(3):205-212. [doi: 10.1089/g4h.2018.0087] [Medline: 31045446]
28. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot Study Assessing the Influence of Skin Type on the Heart Rate Measurements Obtained by Photoplethysmography with the Apple Watch. J Med Syst 2019 May 22;43(7):195. [doi: 10.1007/s10916-019-1325-2] [Medline: 31119387]
29. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 2017;12(2):e0171720 [FREE Full text] [doi: 10.1371/journal.pone.0171720] [Medline: 28234979]
30. Nuss KJ, Thomson EA, Courtney JB, Comstock A, Reinwald S, Blake S, et al. Assessment of Accuracy of Overall Energy Expenditure Measurements for the Fitbit Charge HR 2 and Apple Watch. Am J Health Behav 2019 May 01;43(3):498-505. [doi: 10.5993/ajhb.43.3.5]

31. Zhang P, Godin SD, Owens MV. Measuring the validity and reliability of the Apple Watch as a physical activity monitor. J Sports Med Phys Fitness 2019 May;59(5). [doi: 10.23736/s0022-4707.18.08339-1]
32. Fokkema T, Kooiman T, Krijnen W, Van DS, De GM. Reliability and validity of ten consumer activity trackers depend on walking speed. Medicine and Science in Sport and Exercise 2017;49(4):793-800. [doi: 10.1249/mss.0000000000001146]
33. Breteler MJ, Janssen JH, Spiering W, Kalkman CJ, van Solinge WW, Dohmen DA. Measuring Free-Living Physical Activity With Three Commercially Available Activity Monitors for Telemonitoring Purposes: Validation Study. JMIR Form Res 2019 Apr 24;3(2):e11489 [FREE Full text] [doi: 10.2196/11489] [Medline: 31017587]
34. Gaz DV, Rieck TM, Peterson NW, Ferguson JA, Schroeder DR, Dunfee HA, et al. Determining the Validity and Accuracy of Multiple Activity-Tracking Devices in Controlled and Free-Walking Conditions. Am J Health Promot 2018 Nov;32(8):1671-1678. [doi: 10.1177/0890117118763273] [Medline: 29558811]
35. Modave F, Guo Y, Bian J, Gurka MJ, Parish A, Smith MD, et al. Mobile Device Accuracy for Step Counting Across Age Groups. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e88 [FREE Full text] [doi: 10.2196/mhealth.7870] [Medline: 28659255]
36. Sears T, Alvalos E, Lawson S, McAlister I, Bunn J. Wrist-Worn Physical Activity Trackers Tend To Underestimate Steps During Walking. International Journal of Exercise Science 2017;10(5):764-773 [FREE Full text]
37. Veerabhadrappa P, Moran MD, Renninger MD, Rhudy MB, Dreisbach SB, Gift KM. Tracking Steps on Apple Watch at Different Walking Speeds. J Gen Intern Med 2018 Jun;33(6):795-796 [FREE Full text] [doi: 10.1007/s11606-018-4332-y] [Medline: 29633143]
38. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of Heart Rate Watches: Implications for Weight Management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
39. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of Apple Watch Measurements for Heart Rate and Energy Expenditure in Patients With Cardiovascular Disease: Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e11889 [FREE Full text] [doi: 10.2196/11889] [Medline: 30888332]
40. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov A, et al. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
41. Lee M, Lee H, Park S. Accuracy of swimming wearable watches for estimating energy expenditure. International Journal of Applied Sport Science 2018;30(1):80-90. [doi: 10.24985/ijass.2018.30.1.80]
42. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
43. Boudreaux B, Hebert E, Hollander D, Williams B, Cormier C, Naquin M, et al. Validity of Wearable Activity Monitors during Cycling and Resistance Exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
44. Hwang J, Kim J, Choi K, Cho MS, Nam G, Kim Y. Assessing Accuracy of Wrist-Worn Wearable Devices in Measurement of Paroxysmal Supraventricular Tachycardia Heart Rate. Korean Circ J 2019 May;49(5):437-445 [FREE Full text] [doi: 10.4070/kcj.2018.0323] [Medline: 30808083]
45. Brewer W, Swanson BT, Ortiz A. Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 2017;3(1):e000254 [FREE Full text] [doi: 10.1136/bmjsem-2017-000254] [Medline: 29018543]
46. Hergenroeder AL, Barone Gibbs B, Kotlarczyk MP, Perera S, Kowalsky RJ, Brach JS. Accuracy and Acceptability of Commercial-Grade Physical Activity Monitors in Older Adults. J Aging Phys Act 2019 Apr 01;27(2):222-229 [FREE Full text] [doi: 10.1123/japa.2018-0036] [Medline: 30117355]
47. Husted HM, Llewellyn TL. The Accuracy of Pedometers in Measuring Walking Steps on a Treadmill in College Students. Int J Exerc Sci 2017;10(1):146-153 [FREE Full text] [Medline: 28479955]
48. DeShaw K, Ellingson L, Bai Y, Lansing J, Perez M, Welk G. Methods for Activity Monitor Validation Studies: An Example With the Fitbit Charge. Journal for the Measurement of Physical Behaviour 2018;1(3):130-135. [doi: 10.1123/jmpb.2018-0017]
49. Hargens TA, Deyarmin KN, Snyder KM, Mihalik AG, Sharpe LE. Comparison of wrist-worn and hip-worn activity monitors under free living conditions. J Med Eng Technol 2017 Apr;41(3):200-207. [doi: 10.1080/03091902.2016.1271046] [Medline: 28078908]
50. Smith JD, Guerra G, Burkholder BG. The validity and accuracy of wrist-worn activity monitors in lower-limb prosthesis users. Disabil Rehabil 2019 Apr 12:1-7. [doi: 10.1080/09638288.2019.1587792] [Medline: 30978125]
51. Toth LP, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-Recorded Validation of Wearable Step Counters under Free-living Conditions. Medicine & Science in Sports & Exercise 2018;50(6):1315-1322. [doi: 10.1249/mss.0000000000001569]
52. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of Different Activity Monitors to Count Steps in an Inpatient Rehabilitation Setting. Phys Ther 2017 May 01;97(5):581-588. [doi: 10.1093/ptj/pzx010] [Medline: 28339904]

53. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-Validity of Commercially Available Physical Activity Tracker to Estimate Step Count, Covered Distance and Energy Expenditure during Sports Conditions. Front Physiol 2017;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
54. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity Trackers. Ann Intern Med 2017 Apr 18;166(8):610-612 [FREE Full text] [doi: 10.7326/L16-0353] [Medline: 28395305]
55. Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. J Med Eng Technol 2016 Aug;40(6):325-329. [doi: 10.1080/03091902.2016.1197978] [Medline: 27280592]
56. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The Accuracy of Smart Devices for Measuring Physical Activity in Daily Life: Validation Study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e10972 [FREE Full text] [doi: 10.2196/10972] [Medline: 30545810]
57. Poojary J, Arora E, Britto A, Polen Z, Arena R, Babu AS. Validity of Mobile-Based Technology vs Direct Observation in Measuring Number of Steps and Distance Walked in 6 Minutes. Mayo Clin Proc 2018 Dec;93(12):1873-1874. [doi: 10.1016/j.mayocp.2018.09.003] [Medline: 30522597]
58. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. PLoS One 2019;14(5):e0216891 [FREE Full text] [doi: 10.1371/journal.pone.0216891] [Medline: 31112585]
59. Collins JE, Yang HY, Trentadue TP, Gong Y, Losina E. Validation of the Fitbit Charge 2 compared to the ActiGraph GT3X+ in older adults with knee osteoarthritis in free-living conditions. PLoS One 2019;14(1):e0211231 [FREE Full text] [doi: 10.1371/journal.pone.0211231] [Medline: 30699159]
60. Keating XD, Liu J, Liu X, Shangguan R, Guan J, Chen L. Validity of Fitbit Charge 2 in Controlled College Physical Education Settings. ICHPER-SD Journal of Research 2012;9(2):28-35 [FREE Full text]
61. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018;13(2):e0192691 [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
62. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. Chronobiol Int 2019 Jul;36(7):927-933. [doi: 10.1080/07420528.2019.1596947] [Medline: 30990098]
63. Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, et al. Accuracy of Wrist-Worn Activity Monitors During Common Daily Physical Activities and Types of Structured Exercise: Evaluation Study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e10338 [FREE Full text] [doi: 10.2196/10338] [Medline: 30530451]
64. Dondzila C, Lewis C, Lopez J, Parker T. Congruent Accuracy of Wrist-worn Activity Trackers during Controlled and Free-living Conditions. International Journal of Exercise Science 2018;11(7):575-584 [FREE Full text]
65. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. Gait Posture 2018 Jun;63:104-108. [doi: 10.1016/j.gaitpost.2018.04.034] [Medline: 29729611]
66. Montoye AH, Mitrzyk JR, Molesky MJ. Comparative Accuracy of a Wrist-Worn Activity Tracker and a Smart Shirt for Physical Activity Assessment. Measurement in Physical Education and Exercise Science 2017 Jun 08;21(4):201-211. [doi: 10.1080/1091367X.2017.1331166]
67. Jo E, Lewis K, Directo D, Kim M, Dolezal B. Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking. J Sports Sci Med 2016 Sep;15(3):540-547 [FREE Full text] [Medline: 27803634]
68. Kroll RR, Boyd JG, Maslove DM. Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
69. Pelizzo G, Guddo A, Puglisi A, De Silvestri A, Comparato C, Valenza M, et al. Accuracy of a Wrist-Worn Heart Rate Sensing Device during Elective Pediatric Surgical Procedures. Children (Basel) 2018 Mar 08;5(3) [FREE Full text] [doi: 10.3390/children5030038] [Medline: 29518020]
70. Powierza CS, Clark MD, Hughes JM, Carneiro KA, Mihalik JP. Validation Of A Self-monitoring Tool For Use In Post-concussion Syndrome Therapy. Medicine & Science in Sports & Exercise 2016;48:798. [doi: 10.1249/01.mss.0000487390.22076.d1]
71. Stahl SE, An H, Dinkel DM, Noble JM, Lee J. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):e000106 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
72. Burton E, Hill KD, Lautenschlager NT, Thøgersen-Ntoumani C, Lewin G, Boyle E, et al. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. BMC Geriatr 2018 May 03;18(1):103 [FREE Full text] [doi: 10.1186/s12877-018-0793-4] [Medline: 29724191]
73. Morris CE, Wessel PA, Tinius RA, Schafer MA, Maples JM. Validity of Activity Trackers in Estimating Energy Expenditure During High-Intensity Functional Training. Res Q Exerc Sport 2019 Sep;90(3):377-384. [doi: 10.1080/02701367.2019.1603989] [Medline: 31046630]
74. Farina N, Lowry RG. The Validity of Consumer-Level Activity Monitors in Healthy Older Adults in Free-Living Conditions. J Aging Phys Act 2018 Jan 01;26(1):128-135. [doi: 10.1123/japa.2016-0344] [Medline: 28595019]

75. Rozanski GM, Aqui A, Sivakumaran S, Mansfield A. Consumer Wearable Devices for Activity Monitoring Among Individuals After a Stroke: A Prospective Comparison. JMIR Cardio 2018 Jan 04;2(1):e1 [FREE Full text] [doi: 10.2196/cardio.8199] [Medline: 31758760]
76. Voss C, Gardner RF, Dean PH, Harris KC. Validity of Commercial Activity Trackers in Children With Congenital Heart Disease. Can J Cardiol 2017 Jun;33(6):799-805. [doi: 10.1016/j.cjca.2016.11.024] [Medline: 28347581]
77. Brazendale K, Decker L, Hunt ET, Perry MW, Brazendale AB, Weaver RG, et al. Validity and Wearability of Consumer-based Fitness Trackers in Free-living Children. Int J Exerc Sci 2019;12(5):471-482 [FREE Full text] [Medline: 30899354]
78. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit Charge HR Wireless Heart Rate Monitor: Validation Study Conducted Under Free-Living Conditions. JMIR Mhealth Uhealth 2017 Oct 20;5(10):e157 [FREE Full text] [doi: 10.2196/mhealth.8233] [Medline: 29055881]
79. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent Validity of Wearable Activity Trackers Under Free-Living Conditions. Journal of Strength and Conditioning Research 2017;31(4):1097-1106. [doi: 10.1519/jsc.0000000000001571]
80. Chow J, Thom J, Wewege M, Ward R, Parmenter B. Accuracy of step count measured by physical activity monitors: The effect of gait speed and anatomical placement site. Gait Posture 2017 Sep;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
81. Tam KM, Cheung SY. Validation of Electronic Activity Monitor Devices During Treadmill Walking. Telemed J E Health 2018 Oct;24(10):782-789. [doi: 10.1089/tmj.2017.0263] [Medline: 29364065]
82. Tophøj KH, Petersen MG, Sæbye C, Baad-Hansen T, Wagner S. Validity and Reliability Evaluation of Four Commercial Activity Trackers' Step Counting Performance. Telemed J E Health 2018 Sep;24(9):669-677. [doi: 10.1089/tmj.2017.0264] [Medline: 29303680]
83. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. Med Eng Phys 2014 Jun;36(6):659-669 [FREE Full text] [doi: 10.1016/j.medengphy.2014.02.006] [Medline: 24656871]
84. Phillips LJ, Petroski GF, Markis NE. A Comparison of Accelerometer Accuracy in Older Adults. Res Gerontol Nurs 2015;8(5):213-219. [doi: 10.3928/19404921-20150429-03] [Medline: 25942386]
85. Adam Noah J, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. J Med Eng Technol 2013 Oct;37(7):456-462. [doi: 10.3109/03091902.2013.831135] [Medline: 24007317]
86. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. Med Sci Sports Exerc 2013 Nov;45(11):2105-2112 [FREE Full text] [doi: 10.1249/MSS.0b013e318299d2eb] [Medline: 23669877]
87. Sasaki JE, Hickey A, Mavilia M, Tedesco J, John D, Kozey Keadle S, et al. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. J Phys Act Health 2015 Feb;12(2):149-154. [doi: 10.1123/jpah.2012-0495] [Medline: 24770438]
88. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities. JMIR Mhealth Uhealth 2017 Aug 11;5(8):e106. [doi: 10.2196/mhealth.6321]
89. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
90. Floegel TA, Florez-Pregonero A, Hekler EB, Buman MP. Validation of Consumer-Based Hip and Wrist Activity Monitors in Older Adults With Varied Ambulatory Abilities. J Gerontol A Biol Sci Med Sci 2017 Feb;72(2):229-236 [FREE Full text] [doi: 10.1093/gerona/glw098] [Medline: 27257217]
91. Gilmore SJ, Davidson M, Hahne AJ, McClelland JA. The validity of using activity monitors to detect step count after lumbar fusion surgery. Disabil Rehabil 2020 Mar;42(6):863-868. [doi: 10.1080/09638288.2018.1509140] [Medline: 30326754]
92. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait Posture 2016 Jul;48:36-41. [doi: 10.1016/j.gaitpost.2016.04.025] [Medline: 27477705]
93. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 2018 Jul;52(13):844-850. [doi: 10.1136/bjsports-2016-096990] [Medline: 28483930]
94. Jones D, Crossley K, Dascombe B, Hart HF, Kemp J. Validity and Reliability of the Fitbit FlexTM and Actigraph Gt3X+ At Jogging and Running Speeds. Intl J Sports Phys Ther 2018 Aug;13(5):860-870. [doi: 10.26603/ijspt20180860]
95. Kendall B, Bellovary B, Gothe NP. Validity of wearable activity monitors for tracking steps and estimating energy expenditure during a graded maximal treadmill test. J Sports Sci 2019 Jan;37(1):42-49. [doi: 10.1080/02640414.2018.1481723] [Medline: 29863968]
96. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]

97. Montes J, Young J, Tandy R, Navalta J. Fitbit Flex Energy Expenditure and Step Count Evaluation. Journal Exercise Physiology Online 2017;20(5):134 [FREE Full text]
98. Sala DA, Grissom HE, Delsole EM, Chu ML, Godfried DH, Bhattacharyya S, et al. Measuring ambulation with wrist-based and hip-based activity trackers for children with cerebral palsy. Dev Med Child Neurol 2019 Nov;61(11):1309-1313. [doi: 10.1111/dmcn.14207] [Medline: 30883727]
99. Schmal H, Holsgaard-Larsen A, Izadpanah K, Brønd JC, Madsen CF, Lauritsen J. Validation of Activity Tracking Procedures in Elderly Patients after Operative Treatment of Proximal Femur Fractures. Rehabil Res Pract 2018;2018:3521271 [FREE Full text] [doi: 10.1155/2018/3521271] [Medline: 30018824]
100. Balto JM, Kinnett-Hopkins DL, Motl RW. Accuracy and precision of smartphone applications and commercially available motion sensors in multiple sclerosis. Mult Scler J Exp Transl Clin 2016;2:2055217316634754 [FREE Full text] [doi: 10.1177/2055217316634754] [Medline: 28607720]
101. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and Reliability of Fitbit Flex for Step Count, Moderate to Vigorous Physical Activity and Activity Energy Expenditure. PLoS One 2016;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
102. Ummels D, Beekman E, Theunissen K, Braun S, Beurskens AJ. Counting Steps in Activities of Daily Living in People With a Chronic Disease Using Nine Commercially Available Fitness Trackers: Cross-Sectional Validity Study. JMIR Mhealth Uhealth 2018 Apr 02;6(4):e70 [FREE Full text] [doi: 10.2196/mhealth.8524] [Medline: 29610110]
103. Nelson M, Kaminsky L, Dickin D, Montoye A. Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types. Med Sci Sports Exerc 2016 Aug;48(8):1619-1628. [doi: 10.1249/MSS.0000000000000933] [Medline: 27015387]
104. Murakami H, Kawakami R, Nakae S, Nakata Y, Ishikawa-Takata K, Tanaka S, et al. Accuracy of Wearable Devices for Estimating Total Energy Expenditure: Comparison With Metabolic Chamber and Doubly Labeled Water Method. JAMA Intern Med 2016 May 01;176(5):702-703. [doi: 10.1001/jamainternmed.2016.0152] [Medline: 26999758]
105. Bai Y, Welk GJ, Nam YH, Lee JA, Lee J, Kim Y, et al. Comparison of Consumer and Research Monitors under Semistructured Settings. Medicine & Science in Sports & Exercise 2016;48(1):151-158. [doi: 10.1249/mss.0000000000000727]
106. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Sep;23(14):1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
107. Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
108. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical Activity Assessment Between Consumer- and Research-Grade Accelerometers: A Comparative Study in Free-Living Conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
109. Block VJ, Lizée A, Crabtree-Hartman E, Bevan CJ, Graves JS, Bove R, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. J Neurol 2017 Feb;264(2):316-326 [FREE Full text] [doi: 10.1007/s00415-016-8334-6] [Medline: 27896433]
110. Reid RE, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jun;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]
111. Rowe VT, Neville M. Measuring Reliability of Movement With Accelerometry: Fitbit Versus ActiGraph. Am J Occup Ther 2019;73(2):7302205150p1-7302205150p6. [doi: 10.5014/ajot.2019.030692] [Medline: 30915976]
112. St-Laurent A, Mony MM, Mathieu M, Ruchat SM. Validation of the Fitbit Zip and Fitbit Flex with pregnant women in free-living conditions. J Med Eng Technol 2018 May;42(4):259-264. [doi: 10.1080/03091902.2018.1472822] [Medline: 30198806]
113. Clevenger KA, Molesky MJ, Vusich J, Montoye AH. Free-Living Comparison of Physical Activity and Sleep Data from Fitbit Activity Trackers Worn on the Dominant and Nondominant Wrists. Measurement in Physical Education and Exercise Science 2019 Feb 12;23(2):194-204. [doi: 10.1080/1091367x.2019.1577737]
114. Chen MD, Kuo C, Pellegrini C, Hsu M. Accuracy of Wristband Activity Monitors during Ambulation and Activities. Med Sci Sports Exerc 2016 Oct;48(10):1942-1949. [doi: 10.1249/MSS.0000000000000984] [Medline: 27183123]
115. Daligadu J, Pollock CL, Carlaw K, Chin M, Haynes A, Thevaraajah Kopal T, et al. Validation of the Fitbit Flex in an Acute Post-Cardiac Surgery Patient Population. Physiother Can 2018;70(4):314-320 [FREE Full text] [doi: 10.3138/ptc.2017-34] [Medline: 30745716]
116. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit®: An accurate and reliable device for wireless physical activity tracking. Int J Cardiol 2015 Apr 15;185:138-140 [FREE Full text] [doi: 10.1016/j.ijcard.2015.03.038] [Medline: 25795203]

117. Diaz KM, Krupka DJ, Chang MJ, Shaffer JA, Ma Y, Goldsmith J, et al. Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site. BMC Res Notes 2016 Apr 12;9:213 [FREE Full text] [doi: 10.1186/s13104-016-2020-8] [Medline: 27068022]
118. Battenberg AK, Donohoe S, Robertson N, Schmalzried TP. The accuracy of personal activity monitoring devices. Seminars in Arthroplasty 2017 Jun;28(2):71-75. [doi: 10.1053/j.sart.2017.07.006]
119. Prieto-Centurion V, Bracken N, Norwick L, Zaidi F, Mutso AA, Morken V, et al. Can Commercially Available Pedometers Be Used For Physical Activity Monitoring In Patients With COPD Following Exacerbations? Chronic Obstr Pulm Dis 2016;3(3):636-642 [FREE Full text] [doi: 10.15326/jcopdf.3.3.2015.0164] [Medline: 27547817]
120. Arch ES, Sions JM, Horne J, Bodt BA. Step count accuracy of StepWatch and FitBit One™ among individuals with a unilateral transtibial amputation. Prosthet Orthot Int 2018 Oct;42(5):518-526. [doi: 10.1177/0309364618767138] [Medline: 29623810]
121. Klassen T, Simpson L, Lim S, Louie D, Parappilly B, Sakakibara B, et al. "Stepping Up" Activity Poststroke: Ankle-Positioned Accelerometer Can Accurately Record Steps During Slow Walking. Phys Ther 2016 Mar;96(3):355-360 [FREE Full text] [doi: 10.2522/ptj.20140611] [Medline: 26251478]
122. O'Connell S, ÓLaighin G, Kelly L, Murphy E, Beirne S, Burke N, et al. These Shoes Are Made for Walking: Sensitivity Performance Evaluation of Commercial Activity Monitors under the Expected Conditions and Circumstances Required to Achieve the International Daily Step Goal of 10,000 Steps. PLoS One 2016;11(5):e0154956 [FREE Full text] [doi: 10.1371/journal.pone.0154956] [Medline: 27167121]
123. O'Connell S, ÓLaighin G, Quinlan LR. When a Step Is Not a Step! Specificity Analysis of Five Physical Activity Monitors. PLoS One 2017;12(1):e0169616 [FREE Full text] [doi: 10.1371/journal.pone.0169616] [Medline: 28085918]
124. Simpson L, Eng J, Klassen T, Lim S, Louie D, Parappilly B, et al. Capturing step counts at slow walking speeds in older adults: comparison of ankle and waist placement of measuring device. J Rehabil Med 2015 Oct 05;47(9):830-835 [FREE Full text] [doi: 10.2340/16501977-1993] [Medline: 26181670]
125. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. J Sci Med Sport 2014 Sep;17(5):496-500. [doi: 10.1016/j.jsams.2013.10.241] [Medline: 24268570]
126. Klassen TD, Semrau JA, Dukelow SP, Bayley MT, Hill MD, Eng JJ. Consumer-Based Physical Activity Monitor as a Practical Way to Measure Walking Intensity During Inpatient Stroke Rehabilitation. Stroke 2017 Sep;48(9):2614-2617. [doi: 10.1161/strokeaha.117.018175]
127. Leth S, Hansen J, Nielsen O, Dinesen B. Evaluation of Commercial Self-Monitoring Devices for Clinical Purposes: Results from the Future Patient Trial, Phase I. Sensors (Basel) 2017 Jan 22;17(1) [FREE Full text] [doi: 10.3390/s17010211] [Medline: 28117736]
128. Storm FA, Heller BW, Mazzà C. Step detection and activity recognition accuracy of seven physical activity monitors. PLoS One 2015;10(3):e0118723 [FREE Full text] [doi: 10.1371/journal.pone.0118723] [Medline: 25789630]
129. Lee J, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc 2014 Sep;46(9):1840-1848. [doi: 10.1249/MSS.0000000000000287] [Medline: 24777201]
130. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. J Med Eng Technol 2017 Apr;41(3):208-215. [doi: 10.1080/03091902.2016.1253795] [Medline: 27919170]
131. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
132. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. J Med Internet Res 2016 Sep 07;18(9):e239 [FREE Full text] [doi: 10.2196/jmir.5531] [Medline: 27604226]
133. Hamari L, Kullberg T, Ruohonen J, Heinonen O, Díaz-Rodríguez N, Lilius J, et al. Physical activity among children: objective measurements using Fitbit One and ActiGraph. BMC Res Notes 2017 Apr 20;10(1):161 [FREE Full text] [doi: 10.1186/s13104-017-2476-1] [Medline: 28427441]
134. Hui J, Heyden R, Bao T, Accettone N, McBay C, Richardson J, et al. Validity of the Fitbit One for Measuring Activity in Community-Dwelling Stroke Survivors. Physiother Can 2018;70(1):81-89 [FREE Full text] [doi: 10.3138/ptc.2016-40.ep] [Medline: 29434422]
135. Middelweerd A, Van Der Ploeg HP, Van Halteren A, Twisk JW, Brug J, Te Velde SJ. A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals. Med Sci Sports Exerc 2017 Jun;49(6):1270-1279. [doi: 10.1249/MSS.0000000000001225] [Medline: 28511193]
136. Van Blarigan EL, Kenfield SA, Tantum L, Cadmus-Bertram LA, Carroll PR, Chan JM. The Fitbit One Physical Activity Tracker in Men With Prostate Cancer: Validation Study. JMIR Cancer 2017 Apr 18;3(1):e5 [FREE Full text] [doi: 10.2196/cancer.6935] [Medline: 28420602]
137. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]

138. Duclos NC, Aguiar LT, Aissaoui R, Faria CD, Nadeau S, Duclos C. Activity Monitor Placed at the Nonparetic Ankle Is Accurate in Measuring Step Counts During Community Walking in Poststroke Individuals: A Validation Study. PM R 2019 Sep;11(9):963-971. [doi: 10.1002/pmrj.12080] [Medline: 30688026]
139. Bunn JA, Jones C, Oliviera A, Webster MJ. Assessment of step accuracy using the Consumer Technology Association standard. J Sports Sci 2019 Feb 29;37(3):244-248. [doi: 10.1080/02640414.2018.1491941] [Medline: 29958058]
140. Navalta J, Montes J, Bodell N, Aguilar C, Lujan A, Guzman G, et al. Wearable Device Validity in Determining Step Count During Hiking and Trail Running. Journal for the Measurement of Physical Behaviour 2018;1(2):86-93. [doi: 10.1123/jmpb.2018-0010]
141. Wendel N, Macpherson C, Webber K, Hendron K, DeAngelis T, Colon-Semenza C, et al. Accuracy of Activity Trackers in Parkinson Disease: Should We Prescribe Them? Phys Ther 2018 Aug 01;98(8):705-714. [doi: 10.1093/ptj/pzy054] [Medline: 29718452]
142. Thiebaud RS, Funk MD, Patton JC, Massey BL, Shay TE, Schmidt MG, et al. Validity of wrist-worn consumer products to measure heart rate and energy expenditure. Digit Health 2018 Apr 13;4:2055207618770322 [FREE Full text] [doi: 10.1177/2055207618770322] [Medline: 29942628]
143. Pribyslavska V, Caputo JL, Coons JM, Barry VW. Impact of EPOC adjustment on estimation of energy expenditure using activity monitors. J Med Eng Technol 2018 May 18;42(4):265-273. [doi: 10.1080/03091902.2018.1472823] [Medline: 29911930]
144. Fulk G, Combs S, Danks K, Nirider C, Raja B, Reisman D. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. Phys Ther 2014 Feb;94(2):222-229. [doi: 10.2522/ptj.20120525] [Medline: 24052577]
145. Park W, Lee VJ, Ku B, Tanaka H. Effect of walking speed and placement position interactions in determining the accuracy of various newer pedometers. Journal of Exercise Science & Fitness 2014 Jun;12(1):31-37. [doi: 10.1016/j.jesf.2014.01.003]
146. Stackpool C, Porcari J, Mikat R, Gillette C, Foster C. The Accuracy of Various Activity Trackers in Estimating Steps Taken and Energy Expenditure. J Fit Res 2014;3(3):32-48 [FREE Full text]
147. Wong CK, Mentis HM, Kuber R. The bit doesn't fit: Evaluation of a commercial activity-tracker at slower walking speeds. Gait Posture 2018 Jan;59:177-181. [doi: 10.1016/j.gaitpost.2017.10.010] [Medline: 29049964]
148. Gusmer R, Bosch T, Watkins A, Ostrem J, Dengel D. Comparison of FitBit® Ultra to ActiGraph™ GT1M for Assessment of Physical Activity in Young Adults During Treadmill Walking. TOSMJ 2014 Apr 04;8(1):11-15. [doi: 10.2174/1874387001408010011]
149. Paul SS, Tiedemann A, Hassett LM, Ramsay E, Kirkham C, Chagpar S, et al. Validity of the activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport Exerc Med 2015;1(1):e000013 [FREE Full text] [doi: 10.1136/bmjsem-2015-000013] [Medline: 27900119]
150. Schaffer SD, Holzapfel SD, Fulk G, Bosch PR. Step count accuracy and reliability of two activity tracking devices in people after stroke. Physiother Theory Pract 2017 Oct 04;33(10):788-796. [doi: 10.1080/09593985.2017.1354412] [Medline: 28777710]
151. Sharp CA, Mackintosh KA, Erjavec M, Pascoe DM, Horne PJ. Validity and reliability of the Fitbit Zip as a measure of preschool children's step count. BMJ Open Sport Exerc Med 2017;3(1):e000272 [FREE Full text] [doi: 10.1136/bmjsem-2017-000272] [Medline: 29081984]
152. Singh AK, Farmer C, Van Den Berg ML, Killington M, Barr CJ. Accuracy of the FitBit at walking speeds and cadences relevant to clinical rehabilitation populations. Disabil Health J 2016 Apr;9(2):320-323. [doi: 10.1016/j.dhjo.2015.10.011] [Medline: 26905972]
153. Appelboom G, Taylor BE, Bruce E, Bassile CC, Malakidis C, Yang A, et al. Mobile Phone-Connected Wearable Motion Sensors to Assess Postoperative Mobilization. JMIR Mhealth Uhealth 2015 Jul 28;3(3):e78 [FREE Full text] [doi: 10.2196/mhealth.3785] [Medline: 26220691]
154. Thorup CB, Andreasen JJ, Sørensen EE, Grønkjær M, Dinesen BI, Hansen J. Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. BMJ Open 2017 Mar 31;7(3):e011742 [FREE Full text] [doi: 10.1136/bmjopen-2016-011742] [Medline: 28363918]
155. Mooses K, Oja M, Reisberg S, Vilo J, Kull M. Validating Fitbit Zip for monitoring physical activity of children in school: a cross-sectional study. BMC Public Health 2018 Jul 11;18(1):858 [FREE Full text] [doi: 10.1186/s12889-018-5752-7] [Medline: 29996797]
156. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. BMC Res Notes 2014 Dec 23;7:952 [FREE Full text] [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]
157. Schneider M, Chau L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. BMC Res Notes 2016 Sep 21;9(1):448 [FREE Full text] [doi: 10.1186/s13104-016-2253-6] [Medline: 27655477]
158. Beevi FH, Miranda J, Pedersen CF, Wagner S. An Evaluation of Commercial Pedometers for Monitoring Slow Walking Speed Populations. Telemed J E Health 2016 May;22(5):441-449. [doi: 10.1089/tmj.2015.0120] [Medline: 26451900]
159. Boolani A, Towler C, LeCours B, Blank H, Larue J, Fulk G. Accuracy of 6 Commercially Available Activity Monitors in Measuring Heart Rate, Caloric Expenditure, Steps Walked, and Distance Traveled. Cardiopulmonary Physical Therapy Journal 2019;30(4):153-161. [doi: 10.1097/cpt.0000000000000107]

160. Chandrasekar A, Hensor EM, Mackie SL, Backhouse MR, Harris E. Preliminary concurrent validity of the Fitbit-Zip and ActiGraph activity monitors for measuring steps in people with polymyalgia rheumatica. Gait Posture 2018 Mar;61:339-345. [doi: 10.1016/j.gaitpost.2018.01.035] [Medline: 29427859]
161. Haegele J, Brian A, Wolf D. Accuracy of the Fitbit Zip for Measuring Steps for Adolescents With Visual Impairments. Adapt Phys Activ Q 2017 Apr;34(2):195-200. [doi: 10.1123/apaq.2016-0055] [Medline: 28556761]
162. Claes J, Buys R, Avila A, Finlay D, Kennedy A, Guldenring D, et al. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. J Med Eng Technol 2017 Aug 04;41(6):480-485. [doi: 10.1080/03091902.2017.1333166] [Medline: 28675070]
163. Støve MP, Haucke E, Nymann ML, Sigurdsson T, Larsen BT. Accuracy of the wearable activity tracker Garmin Forerunner 235 for the assessment of heart rate during rest and activity. J Sports Sci 2019 Apr 17;37(8):895-901. [doi: 10.1080/02640414.2018.1535563] [Medline: 30326780]
164. Mahendran N, Kuys SS, Downie E, Ng P, Brauer SG. Are Accelerometers and GPS Devices Valid, Reliable and Feasible Tools for Measurement of Community Ambulation After Stroke? Brain Impairment 2016 Jun 08;17(2):151-161. [doi: 10.1017/brimp.2016.13]
165. Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA. Accuracy of the vivofit activity tracker. J Med Eng Technol 2016 Aug;40(6):298-306. [doi: 10.1080/03091902.2016.1193238] [Medline: 27266422]
166. Woodman J, Crouter S, Bassett D, Fitzhugh E, Boyer W. Accuracy of Consumer Monitors for Estimating Energy Expenditure and Activity Type. Med Sci Sports Exerc 2017 Feb;49(2):371-377. [doi: 10.1249/MSS.0000000000001090] [Medline: 27580155]
167. Simunek A, Dygryn J, Gaba A, Jakubec L, Stelzer J, Chmelik F. Validity of Garmin Vivofit and Polar Loop for measuring daily step counts in free-living conditions in adults. Acta Gymnica 2016 Sep 30;46(3):129-135. [doi: 10.5507/ag.2016.014]
168. Šimůnek A, Dygrýn J, Jakubec L, Neuls F, Frömel K, Welk G. Validity of Garmin Vívofit 1 and Garmin Vívofit 3 for School-Based Physical Activity Monitoring. Pediatr Exerc Sci 2019 Feb 01;31(1):130-136. [doi: 10.1123/pes.2018-0019] [Medline: 30304974]
169. Ehrler F, Weber C, Lovis C. Influence of Pedometer Position on Pedometer Accuracy at Various Walking Speeds: A Comparative Study. J Med Internet Res 2016 Oct 06;18(10):e268 [FREE Full text] [doi: 10.2196/jmir.5916] [Medline: 27713114]
170. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Sports 2018 Jul;28(7):1818-1827. [doi: 10.1111/sms.13074] [Medline: 29460319]
171. De Ridder R, De Blaiser C. Activity trackers are not valid for step count registration when walking with crutches. Gait Posture 2019 May;70:30-32. [doi: 10.1016/j.gaitpost.2019.02.009] [Medline: 30798092]
172. Boeselt T, Spielmanns M, Nell C, Storre JH, Windisch W, Magerhans L, et al. Validity and Usability of Physical Activity Monitoring in Patients with Chronic Obstructive Pulmonary Disease (COPD). PLoS One 2016;11(6):e0157229 [FREE Full text] [doi: 10.1371/journal.pone.0157229] [Medline: 27305105]
173. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015 Dec 17;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]
174. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of Polar V800 accelerometer. Ann Transl Med 2016 Aug;4(15):278-278 [FREE Full text] [doi: 10.21037/atm.2016.07.16] [Medline: 27570772]
175. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. Int J Med Inform 2017 Jun;102:87-92. [doi: 10.1016/j.ijmedinf.2017.03.008] [Medline: 28495352]
176. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of Wrist-Worn Heart Rate Monitors. JAMA Cardiol 2017 Jan 01;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
177. Roos L, Taube W, Beeler N, Wyss T. Validity of sports watches when estimating energy expenditure during running. BMC Sports Sci Med Rehabil 2017 Dec 20;9(1):22 [FREE Full text] [doi: 10.1186/s13102-017-0089-6] [Medline: 29296281]
178. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
179. Gerrior S, Juan W, Basiotis P. An easy approach to calculating estimated energy requirements. Prev Chronic Dis 2006 Oct;3(4):A129 [FREE Full text] [Medline: 16978504]
180. Open mHealth. URL: https://www.openmhealth.org/ [accessed 2020-08-17]
181. Ralls J. Fitbit to Be Acquired by Google. Fitbit. 2019. URL: https://investor.fitbit.com/press/press-releases/press-release-details/2019/Fitbit-to-Be-Acquired-by-Google/default.aspx [accessed 2019-11-01]
182. Osterloh R. Helping more people with wearables: Google to acquire Fitbit Internet. Google Blog. URL: https://blog.google/products/hardware/agreement-with-fitbit?_ga=2.109995341.918473813.1572613323-1996097189.1566566630 [accessed 2020-08-17]

183. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care 2012 Apr;28(2):138-144. [doi: 10.1017/S0266462312000086] [Medline: 22559755]
184. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol 2002 Feb;31(1):115-123. [doi: 10.1093/ije/31.1.115] [Medline: 11914306]

Abbreviations

API: application program interface
COSMIN: Consensus-Based Standards for the Selection of Health Status Measurement Instruments
MAPE: mean absolute percentage error
MPE: mean percentage error

Edited by G Eysenbach; submitted 30.03.20; peer-reviewed by D Lott, L Becker; comments to author 12.06.20; revised version received 22.06.20; accepted 25.06.20; published 08.09.20.

Please cite as:
Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, Buote R, Van Heerden D, Luan H, Cullen K, Slade L, Taylor NGA Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review
JMIR Mhealth Uhealth 2020;8(9):e18694
URL: http://mhealth.jmir.org/2020/9/e18694/
doi:10.2196/18694
PMID:32897239

Original Paper

Data Imputation and Body Weight Variability Calculation Using Linear and Nonlinear Methods in Data Collected From Digital Smart Scales: Simulation and Validation Study

Jake Turicchi[1], BSc; Ruairi O'Driscoll[1], MSc; Graham Finlayson[1], PhD; Cristiana Duarte[1], PhD; A L Palmeira[2], PhD; Sofus C Larsen[3], PhD; Berit L Heitmann[3,4,5], PhD; R James Stubbs[1], PhD

[1]School of Psychology, The University of Leeds, Leeds, United Kingdom

[2]Faculdade de Motricidade Humana, Universidade de Lisboa, Lisbon, Portugal

[3]Research Unit for Dietary Studies, The Parker Institute, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark

[4]The Boden Institute of Obesity, Nutrition and Eating disorder, University of Sydney, Sydney, Australia

[5]Department of Public Health, Section for General Medicine, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Jake Turicchi, BSc
School of Psychology
The University of Leeds
2 Lifton Place
Leeds, LS2 9JS
United Kingdom
Phone: 44 7718300764
Email: psjt@leeds.ac.uk

Abstract

Background: Body weight variability (BWV) is common in the general population and may act as a risk factor for obesity or diseases. The correct identification of these patterns may have prognostic or predictive value in clinical and research settings. With advancements in technology allowing for the frequent collection of body weight data from electronic smart scales, new opportunities to analyze and identify patterns in body weight data are available.

Objective: This study aims to compare multiple methods of data imputation and BWV calculation using linear and nonlinear approaches

Methods: In total, 50 participants from an ongoing weight loss maintenance study (the NoHoW study) were selected to develop the procedure. We addressed the following aspects of data analysis: cleaning, imputation, detrending, and calculation of total and local BWV. To test imputation, missing data were simulated at random and using real patterns of missingness. A total of 10 imputation strategies were tested. Next, BWV was calculated using linear and nonlinear approaches, and the effects of missing data and data imputation on these estimates were investigated.

Results: Body weight imputation using structural modeling with Kalman smoothing or an exponentially weighted moving average provided the best agreement with observed values (root mean square error range 0.62%-0.64%). Imputation performance decreased with missingness and was similar between random and nonrandom simulations. Errors in BWV estimations from missing simulated data sets were low (2%-7% with 80% missing data or a mean of 67, SD 40.1 available body weights) compared with that of imputation strategies where errors were significantly greater, varying by imputation method.

Conclusions: The decision to impute body weight data depends on the purpose of the analysis. Directions for the best performing imputation methods are provided. For the purpose of estimating BWV, data imputation should not be conducted. Linear and nonlinear methods of estimating BWV provide reasonably accurate estimates under high proportions (80%) of missing data.

(JMIR Mhealth Uhealth 2020;8(9):e17977) doi:10.2196/17977

KEYWORDS

weight variability; weight fluctuation; weight cycling; weight instability; imputation; validation; digital tracking; smart scales; body weight; energy balance

Introduction

Background

Recently, the idea of remote health care monitored through a network of internet-connected devices, termed *The (Medical) Internet of Things* [1-3], has become popular, and in 2020, it is thought that 40% of internet of things–related technology is health related, accounting for US $117 billion [4]. With this information, precision medicine will become the future of health care. Frequently tracked body weight data are likely to become a valuable prognostic tool. We have already seen the incorporation of Wi-Fi–connected smart scales into research environments [5-7] accompanied by an increase in popularity and a decrease in costs among the general public. In weight management interventions, 80% and 60% of successful weight loss maintainers report self-weighing weekly and daily, respectively [8]. Regular self-weighing in research environments using tracking technologies will allow for more accurate recognition of body weight patterns, which are currently not well understood.

Body weight variability (BWV), that is, the variability around the overall trend in body weight, can be quantified from frequent body weight measures. Several recent studies have associated BWV with outcomes such as all-cause mortality [9-11], type 2 diabetes incidence [12], cardiovascular morbidity or mortality [13,14], and cancer [15]. Further indications suggest that BWV may serve as a potential prognostic tool for obesity [16,17] and as a risk factor in patients with heart failure [18]. However, significant heterogeneity exists in the methods used to process body weights and define BWV.

Although body weight is a reliable, valid, and simple metric to measure, its short-term dynamics are not well understood because, until recently, it has been difficult and time-consuming to make frequent longitudinal measures from an objective (ie, not self-reported) source, and previous studies estimating BWV generally use infrequent measurements (eg, every 6-12 months). Limitations in the methodologies used may contribute to the poor replicability of the results drawn from differing studies and populations: (1) definitions used and statistical inferences drawn from longitudinal weight data are extremely heterogeneous, (2) body weight changes are often measured retrospectively (by self-report) and/or infrequently (12 months apart), (3) overall trends in body weight (eg, weight increase or decrease) are often not addressed appropriately and may confound independent effects of BWV, and (4) missing data are often not appropriately addressed. Simple linear approaches to the measurement of BWV (such as root mean square error [RMSE] around the linear trend) are not able to fully differentiate the overall trend from the variability component. New strategies must be developed to improve the estimation of BWV.

Using frequent body weight measurements, few studies have examined weekly [19-21] or seasonal [21-23] patterns in body weight, although no study to our knowledge has estimated total BWV over the long term. In future, tracking technologies will become increasingly popular and accompanied by the acquisition of dense and complex data. Appropriate, validated, and accessible data processing methodologies must be devised to deal with such data. Such protocols have been developed for activity tracking [24-26], although they lack body weight tracking.

Objectives

Recently, we collected body weight data from Wi-Fi–connected smart scales in individuals engaged in a weight loss maintenance trial (the NoHoW trial [27]) over 12 months. Therefore, we aimed to develop and evaluate a statistical protocol for analyzing frequent weight data by outlining an approach to cleaning, imputation, detrending, and estimating BWV using frequent body weight data to better inform future practices and quantify the magnitude of errors potentially associated with BWV estimates.

Methods

Materials and Subjects

For the purpose of this analysis, a subsample of 50 individuals were selected from the 1627 participants in the NoHoW trial. The NoHoW study is a 2×2 randomized controlled trial (RCT) testing the efficacy of an information and communications technology–based toolkit for delivering a weight loss maintenance intervention structured around evidence-based strategies related to self-regulation and emotion regulation in the United Kingdom (Leeds), Denmark (Copenhagen), and Portugal (Lisbon). Full inclusion and exclusion criteria and procedures can be found elsewhere [27]. Individuals who participated in the trial had reported ≥5% body weight loss in the 12 months before recruitment. The trial was registered with the ISRCTN registry (ISRCTN88405328). The study was conducted in accordance with the Helsinki Declaration. Ethical approval was granted by local institutional ethics committees at the University of Leeds (17-0082; February 27, 2017), the University of Lisbon (17/2016; February 20, 2017), and the Capital Region of Denmark (H-16030495; March 8, 2017).

The selection of the 50 participants for this analysis was based on those who had the greatest number of weight measurements in the first 12 months of the trial. Selecting those with the greatest completeness of data allowed for (1) better ability to simulate missingness and test imputation performance and (2) more valid baseline estimation of BWV, which can be used to test the agreement with other estimations (in comparison with missing simulated and imputed data). Although the study was an RCT, the structure of the RCT was not used, and all its arms were collapsed. Only 50 individuals were chosen to limit missingness in the observed data, which increases with sample size. All participants were provided with a Fitbit Aria (Fitbit Inc) body weight scale linked to a personalized Fitbit account, and the data were retrieved via the Fitbit app programming interface to a web-based data hub. The device has been shown previously by others to have excellent agreement with a calibrated research-grade SECA 769 scale [28]. Participants were instructed to weigh themselves at least twice per week for the duration of the trial. The characteristics of the participants are presented in Table 1.

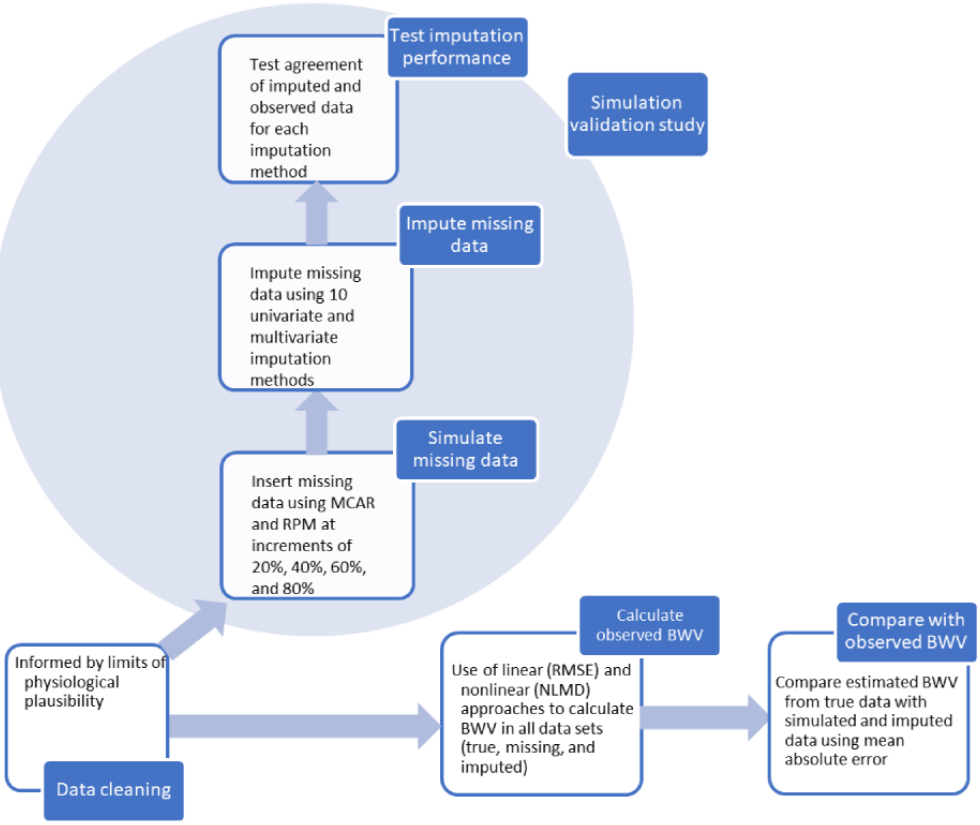
Table 1. Participant characteristics (N=50).

Characteristics	Values
Gender, n (%)	
Male	15 (30)
Female	35 (70)
Age (years), mean (SD)	49.2 (9.3)
Weight (kg), mean (SD)	81.9 (15.4)
BMI (kg/m^2), mean (SD)	29.3 (6.8)
Number of weight measurements, mean (SD)	336.0 (9.1)

Analysis Overview

All statistical analyses were conducted using R version 3.5.1. All statistical codes used can be found in GitHub [29]. A flow diagram of the study is shown in Figure 1. First, we removed outliers based on the limits of physiological plausibility (detailed in Multimedia Appendix 1). Next, we used an amputation and imputation strategy outlined previously [30,31], which involved the simulation of missing data by 2 mechanisms: (1) removal completely at random and (2) removal informed by true patterns of missingness, followed by imputation using univariate and multivariate methods and performance testing using RMSE. Next, we calculated BWV in observed, simulated (ie, inserted missingness), and imputed data sets. This was done to test the accuracy of BWV estimation under conditions of incrementally missing data and when missing data were imputed. BWV was estimated using a commonly used linear approach (RMSE) and a nonlinear approach (nonlinear mean deviation, NLMD) devised for this analysis. Finally, we compared the agreement between BWV estimates from observed weight with those generated by simulated and imputed data sets under different conditions of missingness.

Figure 1. Study flow diagram. Outline of the study detailing the simulation validation study aimed to test imputation performance and calculation of linear and nonlinear body weight variability under conditions of true, missing, and imputed data sets with associated comparisons. BWV: body weight variability; MCAR: missing completely at random; NLMD: nonlinear mean deviation; RMSE: root mean square error; RPM: real patterns of missingness.



Data Cleaning

Data outliers may be present for numerous reasons such as (1) decalibration of electronic scales, (2) inconsistent weighing conditions (eg, clothed vs unclothed or morning vs night), (3) weighing of another person of similar weight (which may register as a rapid weight change on the same Fitbit account), and (4) incorrect manual entry of body weight. We defined the limits of physiological plausibility for weight change over given periods, which can be seen in Multimedia Appendix 1. These limits were informed by substantial weight changes reported during rapid weight loss, such as those achieved by a very low–calorie diet [32,33], and rapid weight gain observed in intentional overfeeding studies [34,35]. It was deemed preferable to remove data based on these plausible limits than risk-removing potentially correct data.

Data Removal

Typically, self-weighing is irregular, and thus, missing data are common. Missing data are generally categorized into missing at random (MAR), missing completely at random (MCAR), or not MAR [36]. Absence of body weight data may have identifiable mechanisms, for example, breaks in self-weighing may be indicative of weight gain [37]; however, these patterns may not be consistent between and within individuals. Data described as MCAR has no mechanism of missingness; however, data that are MAR are not related to the missing data but may be partially explained by the observed data. The data removal processes are described in detail in Multimedia Appendix 2. Briefly, to simulate missing data, we used 2 strategies. First, we inserted data using an MCAR strategy in increments of 20%, 40%, 60%, and 80%. For each of the 50 participants, we simulated 20 data sets per increment of missingness within each participant's data, resulting in 4000 total MCAR-simulated data sets. One potential concern is that missing data in observed data are not entirely MCAR; therefore, MCAR simulation may not be representative of true missingness. To address this, we selected 20 random participants (for each increment of missingness) from our entire NoHoW study sample of 1627 individuals with approximately 20%, 40%, 60%, and 80% missing data and imposed these missing patterns on our 50-participant sample (with a near-complete data), resulting in 4000 simulated data sets with real patterns of missingness (RPM) data. Removing 20%, 40%, 60%, and 80% data left a mean of 255 (SD 54.5), 209 (SD 36.2), 144.8 (SD 50.1), and 67 (SD 40.1) available data points within a year (bearing in mind some data was missing in the original samples).

Data Imputation

Data imputation can be broadly divided into univariate and multivariate approaches. Univariate methods impute missing data based on information gained from a single variable (in this case, a time series [TS] of body weights), whereas multivariate algorithms can be used to infer predictive value from related variables [38] through regression, clustering, or even advanced deep learning techniques. In a remote health care setting, many potentially useful variables for imputing weight data may not be collected (eg, information on psychology and behavior or physiological features), in which case univariate imputation may be necessary. The imputation of univariate TS data lends itself to a limited number of techniques that have been reviewed previously [30].

In total, 7 univariate imputation algorithms and 3 multivariate analyses were run on all missingness-simulated data sets. Univariate methods included (1) linear interpolation; (2) cubic spline interpolation; (3) Stine interpolation; (4) exponentially weighted moving average (EWMA); (5) structural modeling with Kalman smoothing (SMKS); (6) AutoRegressive Integrated Moving Average (ARIMA) state-space representation and Kalman smoothing (ASSRKS), all from the impute TS package [39]; and finally (7) an approach using the Friedman super smoother on nonseasonal data or seasonal decomposition on seasonal data followed by interpolation (TsClean) from the forecast package [40]. Each method is described briefly in Textbox 1. Illustrated examples using a single participant, showing imputation of 40% and 80% missingness by each imputation method, are provided in Multimedia Appendix 3. Multivariate imputation techniques, namely, 2 machine learning techniques (a K-nearest neighbors [KNN] method from the Data Mining with R (DMwR) package [41] and a random forest [RF] method from the MissForest package [42]) and a regression-based technique using predictive means matching (PMM) from the multivariate imputation by the chained equations (MICE) package [43] are described in Textbox 1. To maximize the usability of these methods where further information on participants were not available, we used only the day number and the day of the week as predictive variables for multivariate imputation.

Textbox 1. Description of univariate time series imputation methods used.

Linear interpolation

- This method looks for a straight line that passes between 2 values (Xa and Xb), where the imputed values are bound between Xa and Xb. It has been demonstrated to be efficient when predicting values with constant rate of change [44], however, it tends to smooth data rather than impute variability

Spline interpolation

- This method fits local polynomial functions, which are connected at each end to form a spline, creating a succession of cubic splines over successive intervals of the data [45]. The order of the polynomial can be defined manually. The approach benefits from its nonlinear approach; however, its ability to predict oscillations from univariate data is limited [46]

Stine interpolation

- This is an advanced interpolation method where interpolation occurs based on (1) whether values of the ordinates of the specified points change monotonically and (2) the slopes of the line segments joining the specified points change monotonically. It produces a smoothed imputation known to be robust against sporadic outliers and performs better than spline interpolations, where abrupt changes are observed [47]

Exponentially weighted moving average (EWMA)

- This approach calculates the EWMA by assigning the value of the moving average window, which is user defined; the mean, thereafter, is calculated from equal number of observations on either side of a central missing value. The weighting factors decrease exponentially the greater distance from the missing value

Structural modeling with Kalman smoothing

- This method aims to identify the structure (trend, seasonality, and error) in a time series (TS). Unlike AutoRegressive Integrated Moving Average (ARIMA) state-space approaches where each component is eliminated, these components are used to inform imputation of missing data. Kalman filter and smoothing works in 2 steps to (1) produce estimates of the current state variables, along with their uncertainties, and (2) update estimates using a moving average to give a smoothing effect [48]. The Kalman smoother is given the entire sample and is not locally weighted. The Kalman smoother is robust to disparate observation periods (eg, when observations are made weekly and monthly in one TS) [49]

ARIMA state-space representation and Kalman smoothing

- This method converts the TS to an ARIMA model by decomposing the trend, seasonality, and error through a differencing protocol, resulting in a stationary TS where means and covariances would remain invariant over time [31]. Next, a Kalman smoother is applied as above

TsClean [40]

- This method first assesses evidence of seasonality. If present, a robust seasonal-trend decomposition for seasonal series is conducted followed by linear interpolation. If no seasonality is present, Friedman's super smoother [50] is applied followed by linear interpolation

K-nearest neighbors [41]

- For every observation to be imputed, this algorithm locates *k* closest observations based on the Euclidean distance [51] and computes the weighted average (weighted based on distance) of these *k* observations

Random forest [42]

- This method is an extension of typical classification and regression, which generates predictive models that recursively subdivide the data based on values of the predictor variables. It does not rely on parametric assumptions and can accommodate nonlinear interactions, although it may be prone to overfitting [51]

Predictive means matching [43]

- For each missing entry, this method generates a small set of candidate donors from all complete cases that have predicted values closest to the predicted value for the missing entry. One donor is randomly drawn from the candidates, and the observed value of the donor is taken to replace the missing value. The assumption is the distribution of the missing cell is the same as the observed data of the candidate donors

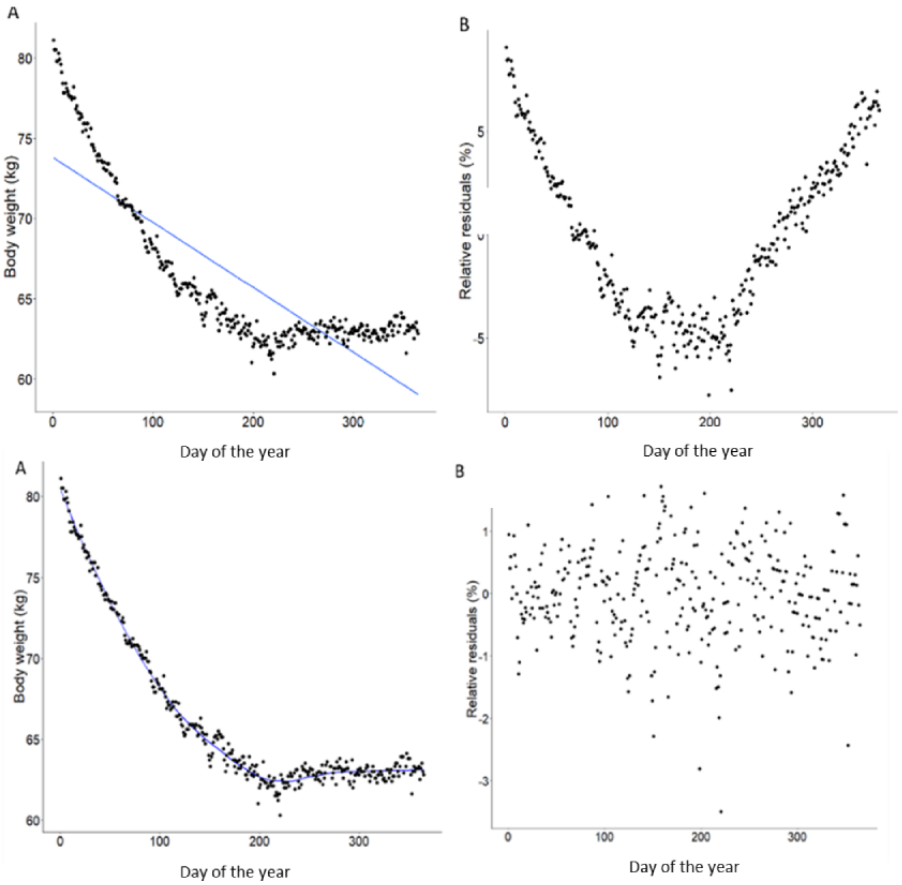
Estimating Body Weight Variability

We estimated BWV using 2 discrete methods in the observed data as well as in all simulated and imputed data sets. These methods are illustrated in Figure 2 for linear (top) and nonlinear (bottom) approaches. First, the RMSE method was used by calculating the relative residual error of the linear relationship between body weight and time (Figure 2). Relative residuals were produced by dividing the centered weight by the observed weight at each time point (Figure 2). This method is commonly used to assess BWV in epidemiological research [16,52-56], although it is limited by the assumption of linearity of body weight change. For example, if an individual displays a curvilinear weight trajectory (such as in Figure 2), then the residuals from a linear trendline will be substantially different from those from a nonlinear trendline. To overcome this, we devised a nonlinear approach detailed below.

First, the series of body weights was detrended for each individual. Detrending is a necessary step in the decomposition of a TS. It can be used to isolate the variability component of the series from the overall trend, resulting in a combination of seasonal patterns (eg, any repetitive patterns including within-week) and random noise. First, a locally estimated scatterplot smoothing (LOESS) regression was fitted to each participant (Figure 2). LOESS regression is a nonlinear, nonparametric smoothing tool. Owing to its nonparametric approach, it does not assume previous specifications about the structure of the data, thus allowing for visual representation of relationships that do not conform to any structure [57]. LOESS regressions were conducted with the *stats* package in R [58]. It employs quadratic polynomial models on a moving collection of data points (termed a *neighborhood*) in a TS [59]. The size of the neighborhood is user defined and referred to as the *span* of the LOESS model, with greater spans creating more smooth trends because of using a wider collection of surrounding data points, whereas shorter spans resulting in closer fitting to the data. The span fits data based on the number of available data; therefore, when fitting the LOESS to data with missingness, the span must be reactive to the number of weight measurements available. To address this, we generated a linear relationship between the span and the number of available data, which resulted in a similar BWV estimation under varying conditions of missingness. Finally, a polynomial order of 2 was used in the model based on the nonlinearity of body weight data, as suggested previously [57].

The detrending process centers body weight around 0. The centered weights were converted to relative centered weights by dividing the centered weight by the observed weight at each time point (Figure 2). This gives an estimate of the relative deviation from the nonlinear trend. BWV was estimated by taking the mean of the centered relative residuals (which act as a proxy of the mean relative deviation from the trend on each day).

Figure 2. Performance summaries of univariate and multivariate imputation. Boxplots of the errors associated with imputation of body weight data collected by smart scales. Data was removed by a missing completely at random algorithm (left plots) and also informed by real patterns of missingness (right plots) in increments of 20%, 40%, 60% and 80%. Imputation was done by 7 univariate methods (top plots) and 3 multivariate methods (bottom plots). Root mean square error was used as the performance metric. ASSRKS: ARIMA state-space representation and Kalman smoothing; EWMA: exponentially weighted moving average; KNN: K-Nearest neighbors; Lin Int: linear interpolation; PMM: predictive means matching; RF: random forest; RMSE: root mean square error; SMKS: structural modelling with Kalman smoothing; Spline int: spline interpolation; Stine int: stine interpolation.



Data Availability

There are legal restrictions on sharing data from this study that contain potentially identifying or sensitive personal information. The restrictions are imposed by the Danish Data Protection Agency Data used in this study will be made available upon request after application to the NoHoW data controller (the James Hutton Institute). The application procedure can be obtained from the James Hutton Institute (DPO@hutton.ac.uk) or David Nutter (david.nutter@bioss.ac.uk).

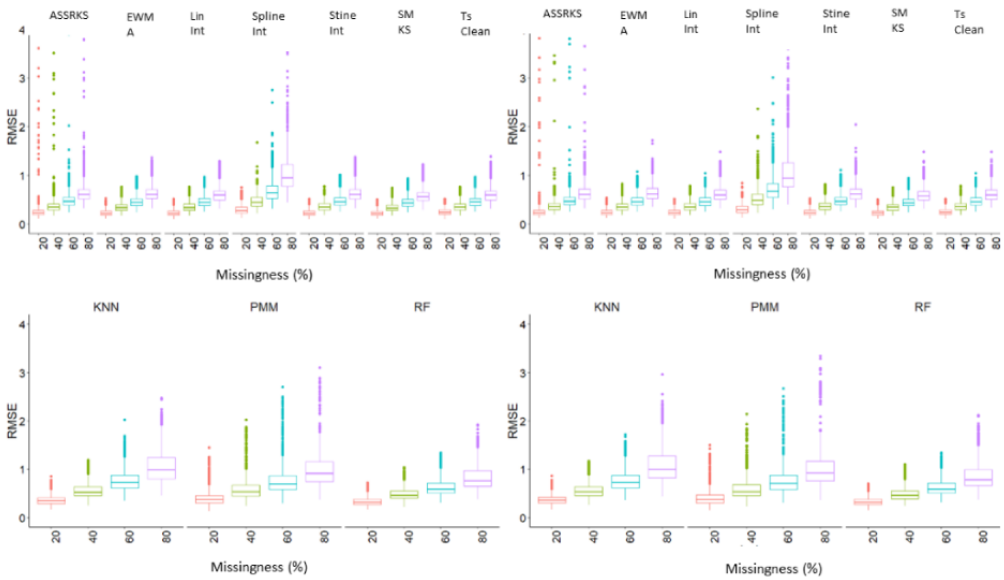
Results

Imputation Performance

All imputation algorithms were run on each simulated data set, generating 28,000 and 12,000 imputed data sets from MCAR and RPM simulations, respectively (4000 imputed data sets per imputation method). The performance of each imputed data set in comparison with the observed weight data was evaluated using the RMSE, which is commonly used for performance evaluation [60]. The RMSE was calculated using the following equation:

The results were grouped by imputation strategy and proportion of missingness. A summary of the performances is illustrated using the RMSE in Figure 3, and the full results are provided in Multimedia Appendix 4. To further test the imputation performance, we used the mean absolute percentage error and mean absolute error, the results of which are shown in Multimedia Appendix 5. The errors increased with greater amounts of missing data. SMKS showed the lowest errors overall, followed by EWMA, linear interpolation, and Stine interpolation, though each of these methods were similar in performance. Machine learning–based methods (RF and KNN) generally performed worse than univariate methods, as did the regression-based multivariate method PMM. The ASSRKS method showed the greatest error, followed by the spline interpolation. Imputation of MCAR-simulated data sets generally showed lower errors than RPM-simulated data sets.

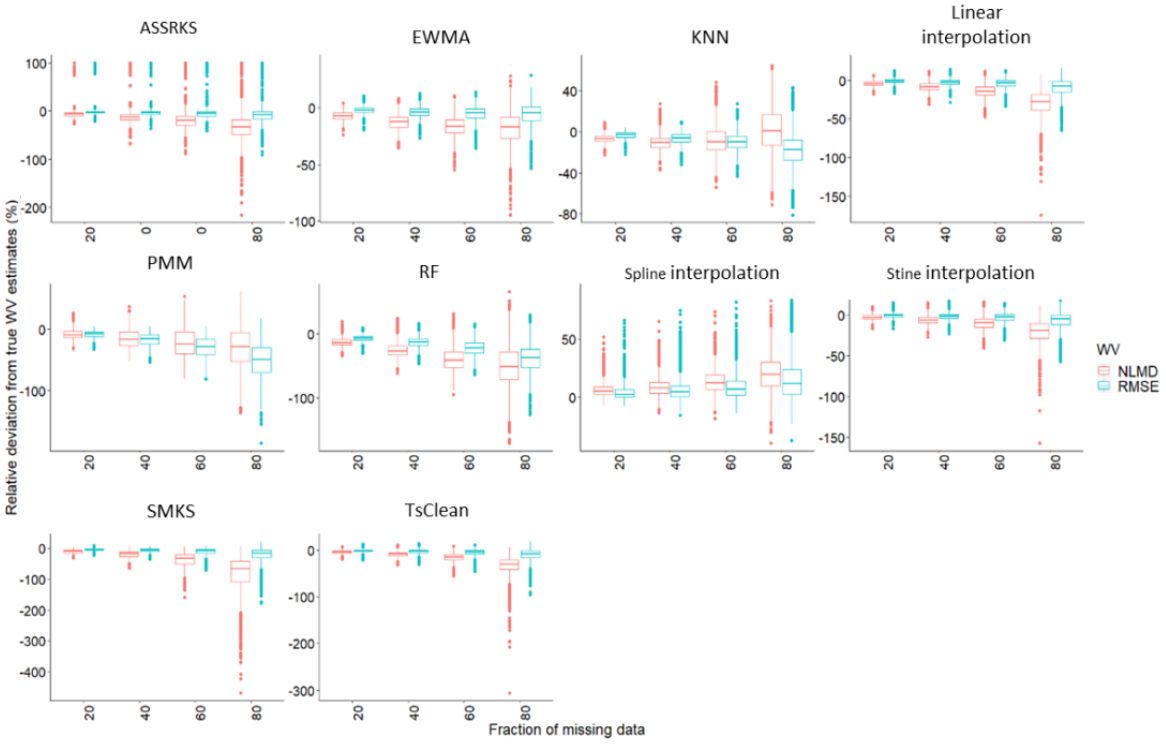
Figure 3. Illustration of linear and non-linear calculation of body weight variability. Scatterplots represent an example of a single participant with a non-linear weight trajectory over 12-months. Figure (A) shows a linear trendline fitted to the data with (B) the trendline subtracted and the associated residuals plotted. Figure (C) shows a non-linear locally estimated scatterplot smoothing regression fitted to the data with (D) the trendline subtracted and the associated residuals plotted. RMSE: root mean square error.



Calculation of BWV

Next, we investigated the agreement between BWV estimations from observed data sets and simulated and imputed data sets for each participant. First, data sets simulated by MCAR and RPM were combined. For simulated data sets (ie, those with missing data), the errors were minimal, reaching an average of 7% (SD 15.4) and 3.2% (SD 19.5) disagreement between the true weight variability (WV) estimates and estimates made on 80% missing data for nonlinear and linear BWV calculation methods, respectively. At 60%, 40%, and 20% of missing data, errors were 2.3% (SD 9.1) and 0.6% (SD 7.3), 1.3% (SD 6.4) and 0.4% (SD 9.8), and 0.4% (SD 6.9) and 0.2% (SD 6.0) for nonlinear and linear WV estimates, respectively, compared with true estimates. The full results can be viewed in Multimedia Appendix 6. When data were imputed, imputation introduced substantial errors in BWV estimates (Figure 4). For most methods, imputation resulted in underestimation of BWV, apart from spline imputation, which overestimated BWV. Biases increased with missingness and were generally greater for NLMD than for RMSE.

Figure 4. Influence of data imputation on linear and non-linear body weight variability estimates. Caption: Boxplots of the relative errors associated with calculation of body weight variability in body weight data collected by smart scales when using 10 different imputation methods imputing data in increments of 20%, 40%, 60%, and 80%. Errors represent the deviation from estimates made from observed data sets. ASSRKS: ARIMA state-space representation and Kalman smoothing; EWMA: exponentially weighted moving average; KNN: K-nearest neighbors; NLMD: nonlinear mean deviation; PMM: predictive means matching; RF: random forest; RMSE: root mean square error; SMKS: structural modeling with Kalman smoothing.



Discussion

In this study, we proposed a method for processing body weight data acquired from electronic smart scales, with both general and specific applications (to BWV). The analysis was produced in response to the increasing use of smart scales in clinical and research environments [24-26]. For the purposes of cleaning, imputation, and detrending, this analysis can inform most researchers dealing with body weight data from smart scales. Furthermore, we provide specific validations on the estimation of BWV using linear and nonlinear approaches and report the errors associated with these estimations when (1) data are missing and (2) data are imputed. We found that SMKS, EWMA, and linear interpolation performed imputation best. These methods are available to researchers through many statistical packages [39]. For the purpose of estimating BWV, we showed that leaving data as missing does not introduce significant bias (only 3%-7% error with >80% data missing), whereas calculating BWV on imputed data causes significant underestimation and should be avoided.

Body Weight Imputation

We considered 7 univariate and 3 multivariate approaches to imputation. As access to further individual-level information (eg, participant characteristics or behavioral patterns and psychological traits) may be unavailable, body weight data collected by smart scales may be treated as univariate, and therefore, the use of more advanced approaches to multivariate imputation such as tree-based models, neural networks, and KNN methods is limited. To test multivariate imputation algorithms, we added day number (ie, day of trial) and day of the week as predictive variables, as these can be automatically collected in free-living environments without any participant burden. Within-week (eg, weekday vs weekend) fluctuations in body weight have been shown previously [19,20] and may have predictive value in imputation. However, we found that these methods, in the current circumstances, did not outperform simple methods such as SMKS or EWMA on MCAR- or RPM-simulated data sets. Indeed, machine learning methods may perform better when trained on large, complete data sets and then applied to missing data; however, in this analysis, we did not have enough complete data sets to train machine learning imputation models, and we chose to limit the variables used in multivariate imputation to improve accessibility. This is the first study to address the issue of missingness and imputation in body weight tracking data; however, several studies have addressed the tracking of physical activity from accelerometers [25,61,62], often using similar simulation and validation approaches with success.

Body Weight Variability Estimation

We proposed a method of estimating BWV using a nonlinear approach, which we termed NLMD. This was devised to address the assumption of linearity associated with RMSE estimations commonly used. Using a nonlinear approach, the trendline is fitted more closely to the data. The result is the ability to identify day-to-day variability or within-week patterns. In contrast, in the case of curvilinear weight trends, RMSE generates large residual errors; this may be more suitable when the aim is to detect larger fluctuations over several months or years. We

found that BWV estimates from data sets with simulated missingness were similar to true estimates, using both RMSE and NLMD methods. Surprisingly, using our current methods, BWV estimates were not greatly different between complete and 80% missing data sets (3.2%, SD 0.2% and 7.0%, SD 0.2% for RMSE and NLMD methods, respectively). However, when these missing data were imputed, substantial biases were introduced largely as underestimations, which increased for each increment of imputed data. As such, although our imputation-validation analysis may inform general imputation of body weight data for numerous other purposes, for the purpose of estimating BWV, we advise that data be left as missing.

To our knowledge, no previous study has examined long-term BWV from electronic smart scales, and only a few studies have modeled frequent weight data using TS methods. A recent study examining the effect of breaks in self-weighing on weight outcomes used a linear mixed model approach using time and weight as fixed predictor and response variables, respectively [37]. However, the use of linear modeling when examining BWV is not sensitive to the often-polynomial features of body weight trajectories. In another study, the authors compared differences in weekday and weekend body weights with longer-term weight changes [20]; however, the data were not detrended. Therefore, weekly weight patterns may potentially be sensitive to overall weight change (particularly in individuals with rapidly changing weight). In a study examining the effect of season on weight patterns across several countries, the authors fitted orthogonal polynomials to the weight data before conducting a detrending process, which may help isolate seasonal patterns from the overall trend (eg, loss or gain) of an individual, showing clear seasonal patterns across the year in different geographical regions [23]. Finally, in a recent study investigating within-week patterns of BWV in 80 adults, the authors took a comprehensive approach by applying nonparametric smoothing techniques (similar to this study) and removed the trend component of the TS using a moving average approach, reporting significant weekly patterns within a week characterized by weekend weight gain and weekday compensation [19]. Recently, we used the present methods to inform the description of weight fluctuation patterns across weeks, years, and holidays [1] and to investigate the associations between BWV and cardiometabolic health outcomes [2].

Strengths and Limitations

This study has several strengths. First, we developed our data processing methods from true rather than simulated data, thus increasing the validity of the analysis. Our simulation-imputation analysis was comprehensive, including the generation of 8000 missingness-simulated data sets in total with varying levels of missingness using both random and real-missingness informed simulations, which resulted in 80,000 imputed data sets produced using 10 univariate and multivariate algorithms. Next, we described and compared both linear and nonlinear approaches to estimating BWV under different conditions of missingness and reported the errors produced in the common case of missing data, which should inform the magnitude of errors expected from missing data estimations in future studies. Some limitations should also be addressed. First, all imputation methods were deterministic, although body weight seems to be a relatively stochastic (ie, randomly determined) process. The resultant effect is that imputation may reduce the variability by attempting to recognize predictive patterns that are not there. We recommend that consideration should be given to whether imputation is necessary. In some analyses, including instances where machine learning algorithms are employed, complete data are a necessity; therefore, imputation is required. Next, we did not have entirely complete data by which to test imputation, although we opted to use real rather than simulated data for external validity.

Conclusions

BWV potentially represents (1) a significant health risk and (2) a prognostic tool that is currently not well understood or well measured. This study evaluated the performance of various imputation methods applied to body weight data and presented a protocol for estimating BWV under varying amounts of missing data. We showed that structural modeling with a Kalman smoother and EWMA performed an imputation most effectively. However, in the case of estimating BWV, the imputations generally produced large underestimations due to the tendency to revert toward the mean. Furthermore, we demonstrated the errors associated with BWV estimates at varying levels of missing data, concluding that errors are small when using both linear and nonlinear methods even under high proportions of missingness. In future, the importance of both frequent measurement of body weight and consistent and appropriate methods of analyzing the data produced should be underlined in the study of BWV.

Acknowledgments

The authors wish to thank all the individuals who were involved in the collection of data at the University of Leeds, the University of Lisbon, and the Capital Region of Denmark. The NoHoW study received funding from the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 643309) in addition to an Engineering and Physical Sciences Research Council (grant no: EP/R511717/1).

Authors' Contributions

JS, AL, SL, and BH were involved in the design of the NoHoW trial. JT, RO, and RS were involved in the conception of this study. JT and RO were involved in the development of the protocol and data analysis. JT was primarily responsible for writing the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

RJS consults for Slimming World via consulting Leeds, a wholly-owned subsidiary of the University of Leeds. All other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Outlier detection limits based on physiological plausibility.
[DOCX File , 13 KB - mhealth_v8i9e17977_app1.docx]

Multimedia Appendix 2
Detailed description of the data removal processes.
[DOCX File , 17 KB - mhealth_v8i9e17977_app2.docx]

Multimedia Appendix 3
Illustrated examples of body weight imputation by all methods.
[DOCX File , 6111 KB - mhealth_v8i9e17977_app3.docx]

Multimedia Appendix 4
Table of imputation performance by root mean square error.
[DOCX File , 16 KB - mhealth_v8i9e17977_app4.docx]

Multimedia Appendix 5
Summary figures of imputation performance by mean absolute percentage error and mean absolute error.
[DOCX File , 109 KB - mhealth_v8i9e17977_app5.docx]

Multimedia Appendix 6
Table of mean errors in body weight variability calculation following simulation and imputation.
[DOCX File , 16 KB - mhealth_v8i9e17977_app6.docx]

References

1. Schreier G. The internet of things for personalized health. Stud Health Technol Inform 2014;200:22-31. [Medline: 24851958]
2. Sheth A, Jaimini U, Yip HY. How will the internet of things enable augmented personalized health? IEEE Intell Syst 2018;33(1):89-97 [FREE Full text] [doi: 10.1109/MIS.2018.012001556] [Medline: 29887765]
3. Wu F, Wu T, Yuce M. An internet-of-things (IoT) network system for connected safety and health monitoring applications. Sensors (Basel) 2018 Dec 21;19(1):21 [FREE Full text] [doi: 10.3390/s19010021] [Medline: 30577646]
4. Dimitrov DV. Medical internet of things and big data in healthcare. Healthc Inform Res 2016 Jul;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
5. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
6. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]
7. Valle CG, Deal AM, Tate DF. Preventing weight gain in African American breast cancer survivors using smart scales and activity trackers: a randomized controlled pilot study. J Cancer Surviv 2017 Feb;11(1):133-148 [FREE Full text] [doi: 10.1007/s11764-016-0571-2] [Medline: 27631874]
8. Vanwormer JJ, French SA, Pereira MA, Welsh EM. The impact of regular self-weighing on weight management: a systematic literature review. Int J Behav Nutr Phys Act 2008 Nov 4;5:54 [FREE Full text] [doi: 10.1186/1479-5868-5-54] [Medline: 18983667]
9. Oh T, Moon J, Choi S, Lim S, Park KS, Cho NH, et al. Body-weight fluctuation and incident diabetes mellitus, cardiovascular disease, and mortality: a 16-year prospective cohort study. J Clin Endocrinol Metab 2019 Mar 1;104(3):639-646. [doi: 10.1210/jc.2018-01239] [Medline: 30500906]
10. Nam GE, Cho KH, Han K, Han B, Cho SJ, Roh YK, et al. Impact of body mass index and body weight variabilities on mortality: a nationwide cohort study. Int J Obes (Lond) 2019 Feb;43(2):412-423. [doi: 10.1038/s41366-018-0079-0] [Medline: 29777238]
11. Kim MK, Han K, Park Y, Kwon H, Kang G, Yoon K, et al. Associations of variability in blood pressure, glucose and cholesterol concentrations, and body mass index with mortality and cardiovascular outcomes in the general population. Circulation 2018 Dec 4;138(23):2627-2637. [doi: 10.1161/CIRCULATIONAHA.118.034978] [Medline: 30571256]

12. Kodama S, Fujihara K, Ishiguro H, Horikawa C, Ohara N, Yachi Y, et al. Unstable bodyweight and incident type 2 diabetes mellitus: a meta-analysis. J Diabetes Investig 2017 Jul;8(4):501-509 [FREE Full text] [doi: 10.1111/jdi.12623] [Medline: 28083921]
13. Aucott LS, Philip S, Avenell A, Afolabi E, Sattar N, Wild S, Scottish Diabetes Research Network Epidemiology Group. Patterns of weight change after the diagnosis of type 2 diabetes in Scotland and their relationship with glycaemic control, mortality and cardiovascular outcomes: a retrospective cohort study. BMJ Open 2016 Jul 26;6(7):e010836 [FREE Full text] [doi: 10.1136/bmjopen-2015-010836] [Medline: 27466237]
14. Bangalore S, Fayyad R, Laskey R, DeMicco DA, Messerli FH, Waters DD. Body-weight fluctuations and outcomes in coronary disease. N Engl J Med 2017 Apr 6;376(14):1332-1340. [doi: 10.1056/nejmoa1606148] [Medline: 28379800]
15. Welti LM, Beavers DP, Caan BJ, Sangi-Haghpeykar H, Vitolins MZ, Beavers KM. Weight fluctuation and cancer risk in postmenopausal women: the women's health initiative. Cancer Epidemiol Biomarkers Prev 2017 May;26(5):779-786 [FREE Full text] [doi: 10.1158/1055-9965.EPI-16-0611] [Medline: 28069684]
16. Lowe M, Feig E, Winter S, Stice E. Short-term variability in body weight predicts long-term weight gain. Am J Clin Nutr 2015 Nov;102(5):995-999 [FREE Full text] [doi: 10.3945/ajcn.115.115402] [Medline: 26354535]
17. Feig EH, Lowe MR. Variability in weight change early in behavioral weight loss treatment: theoretical and clinical implications. Obesity (Silver Spring) 2017 Sep;25(9):1509-1515 [FREE Full text] [doi: 10.1002/oby.21925] [Medline: 28845608]
18. Jimison H, Pavel M. Monitoring of Body Weight for Heart Failure Patients: Variability of Weight and Self-Reporting. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2003 Presented at: IEMBS'03; September 17-21, 2003; Cancun, Mexico. [doi: 10.1109/iembs.2003.1280950]
19. Orsama A, Mattila E, Ermes M, van Gils M, Wansink B, Korhonen I. Weight rhythms: weight increases during weekends and decreases during weekdays. Obes Facts 2014;7(1):36-47 [FREE Full text] [doi: 10.1159/000356147] [Medline: 24504358]
20. Racette SB, Weiss EP, Schechtman KB, Steger-May K, Villareal DT, Obert KA, et al. Influence of weekend lifestyle patterns on body weight. Obesity (Silver Spring) 2008 Aug;16(8):1826-1830 [FREE Full text] [doi: 10.1038/oby.2008.320] [Medline: 18551108]
21. Turicchi J, O'Driscoll R, Horgan G, Duarte C, Palmeira AL, Larsen SC, et al. Weekly, seasonal and holiday body weight fluctuation patterns among individuals engaged in a European multi-centre behavioural weight loss maintenance intervention. PLoS One 2020;15(4):e0232152 [FREE Full text] [doi: 10.1371/journal.pone.0232152] [Medline: 32353079]
22. Fahey MC, Klesges RC, Kocak M, Talcott GW, Krukowski RA. Seasonal fluctuations in weight and self-weighing behavior among adults in a behavioral weight loss intervention. Eat Weight Disord 2020 Aug;25(4):921-928. [doi: 10.1007/s40519-019-00707-7] [Medline: 31093925]
23. Mehrang S, Helander E, Chieh A, Korhonen I. Seasonal Weight Variation Patterns in Seven Countries Located in Northern and Southern Hemispheres. In: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2016 Presented at: EMBC'16; August 16-20, 2016; Orlando, FL, USA. [doi: 10.1109/embc.2016.7591232]
24. Faust L, Wang C, Hachen D, Lizardo O, Chawla NV. Physical activity trend extraction: a framework for extracting moderate-vigorous physical activity trends from wearable fitness tracker data. JMIR Mhealth Uhealth 2019 Mar 12;7(3):e11075 [FREE Full text] [doi: 10.2196/11075] [Medline: 30860488]
25. Catellier DJ, Hannan PJ, Murray DM, Addy CL, Conway TL, Yang S, et al. Imputation of missing data when measuring physical activity by accelerometry. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S555-S562 [FREE Full text] [doi: 10.1249/01.mss.0000185651.59486.4e] [Medline: 16294118]
26. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2:45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
27. Scott SE, Duarte C, Encantado J, Evans EH, Harjumaa M, Heitmann BL, et al. The NoHoW protocol: a multicentre 2×2 factorial randomised controlled trial investigating an evidence-based digital toolkit for weight loss maintenance in European adults. BMJ Open 2019 Sep 30;9(9):e029425 [FREE Full text] [doi: 10.1136/bmjopen-2019-029425] [Medline: 31575569]
28. Shaffer JA, Diaz K, Alcántara C, Edmondson D, Krupka DJ, Chaplin WF, et al. An inexpensive device for monitoring patients' weights via automated hovering. Int J Cardiol 2014 Mar 15;172(2):e263-e264 [FREE Full text] [doi: 10.1016/j.ijcard.2013.12.123] [Medline: 24508493]
29. Turicchi J. BodyWeight Imputation Validation Variability. GitHub. 2020. URL: https://github.com/jaketuricchi/BodyWeight_Imputation_Validation_Variability [accessed 2020-06-13]
30. Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J. Comparison of different methods for univariate time series imputation in R. ArXiv 2015:- epub ahead of print [FREE Full text]
31. Rantou K, Karagrigoriou A, Vonta I. On Imputation Methods in Univariate Time Series. Nonlinear Studies. 2017. URL: http://nonlinearstudies.com/index.php/mesa/article/view/1522 [accessed 2019-06-20]
32. Saris W. Very-low-calorie diets and sustained weight loss. Obes Res 2001 Nov;9(Suppl 4):295S-301S [FREE Full text] [doi: 10.1038/oby.2001.134] [Medline: 11707557]

33. Sellahewa L, Khan C, Lakkunarajah S, Idris I. A systematic review of evidence on the use of very low calorie diets in people with diabetes. Curr Diabetes Rev 2017;13(1):35-46. [doi: 10.2174/1573399812666151005123431] [Medline: 26435354]
34. Dériaz O, Tremblay A, Bouchard C. Non linear weight gain with long term overfeeding in man. Obes Res 1993 May;1(3):179-185 [FREE Full text] [doi: 10.1002/j.1550-8528.1993.tb00609.x] [Medline: 16353352]
35. Leaf A, Antonio J. The effects of overfeeding on body composition: the role of macronutrient composition - a narrative review. Int J Exerc Sci 2017;10(8):1275-1296 [FREE Full text] [Medline: 29399253]
36. Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? Int J Epidemiol 2014 Aug;43(4):1336-1339 [FREE Full text] [doi: 10.1093/ije/dyu080] [Medline: 24706730]
37. Helander EE, Vuorinen A, Wansink B, Korhonen IK. Are breaks in daily self-weighing associated with weight gain? PLoS One 2014;9(11):e113164 [FREE Full text] [doi: 10.1371/journal.pone.0113164] [Medline: 25397613]
38. Spratt M, Carpenter J, Sterne JA, Carlin JB, Heron J, Henderson J, et al. Strategies for multiple imputation in longitudinal studies. Am J Epidemiol 2010 Aug 15;172(4):478-487. [doi: 10.1093/aje/kwq137] [Medline: 20616200]
39. Moritz S. Title Time Series Missing Value Imputation. GitHub. 2018. URL: https://github.com/SteffenMoritz/imputeTS [accessed 2019-06-20]
40. Hyndman R, Athanasopoulos G, Bergmeir C. Package 'Forecast'. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/web/packages/forecast/forecast.pdf [accessed 2019-06-20]
41. Torgo L. Package 'DMwR'. The Comprehensive R Archive Network. 2015. URL: https://cran.r-project.org/web/packages/DMwR/DMwR.pdf [accessed 2020-08-13] [WebCite Cache ID https://cran.r-project.org/web/packages/DMwR/DMwR.pdf]
42. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 1;28(1):112-118. [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
43. Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):-. [doi: 10.18637/jss.v045.i03]
44. Gnauck A. Interpolation and approximation of water quality time series and process identification. Anal Bioanal Chem 2004 Oct;380(3):484-492. [doi: 10.1007/s00216-004-2799-3] [Medline: 15365680]
45. Duan Q, Djidjeli K, Price W, Twizell E. Weighted rational cubic spline interpolation and its application. J Comput Appl Math 2000 May;117(2):121-135. [doi: 10.1016/s0377-0427(99)00336-2]
46. Lepot M, Aubin J, Clemens F. Interpolation in time series: an introductive overview of existing methods, their performance criteria and uncertainty assessment. Water 2017 Oct 17;9(10):796. [doi: 10.3390/w9100796]
47. Perillo GM, Piccolo MC. An interpolation method for estuarine and oceanographic data. Comput Geosci 1991 Jan;17(6):813-820. [doi: 10.1016/0098-3004(91)90062-i]
48. Jalles JT. Structural time series models and the kalman filter: a concise review. SSRN Journal 2009 Jun:- [FREE Full text] [doi: 10.2139/ssrn.1496864]
49. Tusell F. Multiple Imputation of Time Series With an Application to the Construction of Historical Price Indices. Semantic Scholar. 2019. URL: https://pdfs.semanticscholar.org/03d9/7350371f8199a5d7f5d028681e8523b50c56.pdf [accessed 2019-06-26]
50. Friedman J. A Varuble Span Smoother. SLAC National Accelerator Laboratory. 1984. URL: https://www.slac.stanford.edu/pubs/slacpubs/3250/slac-pub-3477.pdf [accessed 2019-06-26]
51. Shah JS, Rai SN, DeFilippis AP, Hill BG, Bhatnagar A, Brock GN. Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies. BMC Bioinformatics 2017 Feb 20;18(1):114 [FREE Full text] [doi: 10.1186/s12859-017-1547-6] [Medline: 28219348]
52. Saito Y, Takahashi O, Arioka H, Kobayashi D. Associations between body fat variability and later onset of cardiovascular disease risk factors. PLoS One 2017;12(4):e0175057 [FREE Full text] [doi: 10.1371/journal.pone.0175057] [Medline: 28369119]
53. Lee J, Kawakubo K, Kobayashi Y, Mori K, Kasihara H, Tamura M. Effects of ten year body weight variability on cardiovascular risk factors in Japanese middle-aged men and women. Int J Obes Relat Metab Disord 2001 Jul;25(7):1063-1067. [doi: 10.1038/sj.ijo.0801633] [Medline: 11443507]
54. Nakanishi N, Nakamura K, Suzuki K, Tatara K. Effects of weight variability on cardiovascular risk factors; a study of nonsmoking Japanese male office workers. Int J Obes Relat Metab Disord 2000 Sep;24(9):1226-1230. [doi: 10.1038/sj.ijo.0801389] [Medline: 11033995]
55. French S, Jeffery R, Folsom A, Williamson D, Byers T. Relation of weight variability and intentionality of weight loss to disease history and health-related variables in a population-based sample of women aged 55-69 years. Am J Epidemiol 1995 Dec 15;142(12):1306-1314. [doi: 10.1093/oxfordjournals.aje.a117598] [Medline: 7503051]
56. Folsom AR, French SA, Zheng W, Baxter JE, Jeffery RW. Weight variability and mortality: the Iowa women's health study. Int J Obes Relat Metab Disord 1996 Aug;20(8):704-709. [Medline: 8856391]
57. Jacoby W. Loess. Electoral Stud 2000 Dec;19(4):577-613. [doi: 10.1016/S0261-3794(99)00028-1]
58. R: The R Stats Package. The R Language - Seminar for Statistics. 2019. URL: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html [accessed 2019-07-08]

59. Siangphoe U, Wheeler DC. Evaluation of the performance of smoothing functions in generalized additive models for spatial variation in disease. Cancer Inform 2015;14(Suppl 2):107-116 [FREE Full text] [doi: 10.4137/CIN.S17300] [Medline: 25983545]
60. Chen C, Twycross J, Garibaldi JM. A new accuracy measure based on bounded relative error for time series forecasting. PLoS One 2017;12(3):e0174202 [FREE Full text] [doi: 10.1371/journal.pone.0174202] [Medline: 28339480]
61. Liu B, Yu M, Graubard BI, Troiano RP, Schenker N. Multiple imputation of completely missing repeated measures data within person from a complex sample: application to accelerometer data in the national health and nutrition examination survey. Stat Med 2016 Dec 10;35(28):5170-5188 [FREE Full text] [doi: 10.1002/sim.7049] [Medline: 27488606]
62. Borghese MM, Borgundvaag E, McIsaac MA, Janssen I. Imputing accelerometer nonwear time in children influences estimates of sedentary time and its associations with cardiometabolic risk. Int J Behav Nutr Phys Act 2019 Jan 17;16(1):7 [FREE Full text] [doi: 10.1186/s12966-019-0770-0] [Medline: 30654817]

Abbreviations

ARIMA: AutoRegressive Integrated Moving Average
ASSRKS: ARIMA state-space representation and Kalman smoothing
BWV: body weight variability
EWMA: exponentially weighted moving average
KNN: K-nearest neighbors
LOESS: locally estimated scatterplot smoothing
MAR: missing at random
MCAR: missing completely at random
NLMD: nonlinear mean deviation
PMM: predictive means matching
RCT: randomized controlled trial
RF: random forest
RMSE: root mean square error
RPM: real patterns of missingness
SMKS: structural modeling with Kalman smoothing
TS: time series
WV: weight variability

Edited by G Eysenbach; submitted 27.01.20; peer-reviewed by V Gupta, R Bajpai; comments to author 12.06.20; accepted 25.06.20; published 11.09.20.

Please cite as:
Turicchi J, O'Driscoll R, Finlayson G, Duarte C, Palmeira AL, Larsen SC, Heitmann BL, Stubbs RJ
Data Imputation and Body Weight Variability Calculation Using Linear and Nonlinear Methods in Data Collected From Digital Smart Scales: Simulation and Validation Study
JMIR Mhealth Uhealth 2020;8(9):e17977
URL: http://mhealth.jmir.org/2020/9/e17977/
doi:10.2196/17977
PMID:32915155

Original Paper

Accuracy of Sedentary Behavior–Triggered Ecological Momentary Assessment for Collecting Contextual Information: Development and Feasibility Study

Marco Giurgiu[1,2], MSc; Christina Niermann[3], PhD; Ulrich Ebner-Priemer[1], Prof Dr; Martina Kanning[3], Prof Dr

[1]Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[3]Department of Sport Science, University of Konstanz, Konstanz, Germany

Corresponding Author:
Marco Giurgiu, MSc
Mental mHealth Lab
Institute of Sports and Sports Science
Karlsruhe Institute of Technology
Hertzstr 16
Karlsruhe, 76131
Germany
Phone: 49 176203557
Email: marco.giurgiu@kit.edu

Abstract

Background: Sedentary behavior has received much attention in the scientific community over the past decade. There is growing evidence that sedentary behavior is negatively associated with physical and mental health. However, an in-depth understanding of the social and environmental context of sedentary behavior is missing. Information about sedentary behavior, such as how everyday sedentary behavior occurs throughout the day (eg, number and length of sedentary bouts), where, when, and with whom it takes place, and what people are doing while being sedentary, is useful to inform the development of interventions aimed at reducing sedentary time. However, examining everyday sedentary behavior requires specific methods.

Objective: The purpose of this paper is (1) to introduce sedentary behavior–triggered Ecological Momentary Assessment (EMA) as a methodological advancement in the field of sedentary behavior research and (2) to examine the accuracy of sedentary behavior–triggered EMA in 3 different studies in healthy adults. Moreover, we compare the accuracy of sedentary behavior–triggered EMA to simulations of random-trigger designs.

Methods: Sedentary behavior–triggered EMA comprises a continuous assessment of sedentary behavior via accelerometers and repeated contextual assessments via electronic diaries (ie, an application on a smartphone). More specifically, the accelerometer analyzes and transfers data regarding body position (a sitting or lying position, or an upright position) via Bluetooth Low Energy (BLE) to a smartphone in real time and triggers the deployment of questionnaires. Each time a participant spends a specified time (eg, 20 minutes) in a sedentary position, the e-diary triggers contextual assessments. To test the accuracy of this method, we calculated a percentage score for all triggered prompts in relation to the total number of bouts that could trigger a prompt.

Results: Based on the accelerometer recordings, 29.3% (5062/17278) of all sedentary bouts were classified as moderate-to-long (20-40 minutes) and long bouts (≥ 41 minutes). On average, the accuracy by participant was 82.77% (3339/4034; SD 21.01%, range 71.00-88.22%) on the study level. Compared to simulations of random prompts (every 120 minutes), the number of triggered prompts was up to 47.9% (n=704) higher through the sedentary behavior–triggered EMA approach. Nearly 40% (799/2001) of all prolonged sedentary bouts (≥ 20 minutes) occurred during work, and in 57% (1140/2001) of all bouts, the participants were not alone.

Conclusions: Sedentary behavior–triggered EMA is an accurate method for collecting contextual information on sedentary behavior in daily life. Given the growing interest in sedentary behavior research, this sophisticated approach offers a real advancement as it can be used to collect social and environmental contextual information or to unravel dynamic associations. Furthermore, it can be modified to develop sedentary behavior–triggered mHealth interventions.

(JMIR Mhealth Uhealth 2020;8(9):e17852) doi:10.2196/17852

KEYWORDS

sedentariness; Ecological Momentary Assessment; accelerometry; mHealth; context

Introduction

Growing Awareness of the Risks of Sedentary Behavior

"Sitting is the new smoking" or "Why a sedentary lifestyle is killing you"—these and similar headlines have received a high level of media attention in recent years. There is growing evidence that sedentary behavior is a behavioral risk factor for human health [1]. In particular, researchers identified that too much sitting is a major risk factor for physical and mental health [2,3]. For example, studies indicated that sedentary behavior is associated with cardiometabolic diseases, diabetes mellitus type 2, and mood disorders [4-6]. Since the amount of evidence has been increasing, countries have started to publish public health guidelines for adults to reduce sedentary time [7,8]. However, currently, there are still uncertainties and divergent views on this behavior [9,10], mainly related to inconsistencies in the definition of sedentary behavior and inaccuracies in the measurement of sedentary behaviors. This paper gives a short overview of sedentary behavior definitions and measurement methods, pointing out the currently recommended ones, and introduces sedentary behavior–triggered Ecological Momentary Assessment (EMA) as an innovative measurement approach for measuring contextual information.

Defining Sedentary Behavior

Several different definitions have evolved over the past decade [11]. From a historical perspective, researchers began by classifying sedentary behavior as physical inactivity. Although sedentary behavior is indeed a form of physical inactivity, the results from physiological studies identified unique mechanisms and characteristics of sedentary behavior and thus suggested that sedentary behavior is an independent behavior, with its own facets and not just the absence of physical activity [12]. Some definitions focused on postural aspects, whereas others focused on energy expenditure without considering postural aspects such as standing versus sitting [11,13], which is questionable since standing may have distinct effects on health outcomes [14-16]. To provide clarity, the Sedentary Behavior Research Network (SBRN) [11] defined sedentary behavior as "any waking behavior characterized by an energy expenditure ≤ 1.5 metabolic equivalents (METs), while in a sitting, reclining, or lying posture." Unlike other definitions, it comprises both components of sedentary behavior (ie, body posture and movement intensity or energy expenditure). Furthermore, this definition clarifies that standing is not a sedentary state. At this time, the definition of the SBRN is internationally accepted, although there are still some ongoing discussions [13]; for example, the threshold of ≤ 1.5 METs is questionable because the amount of energy expended during sitting does exceed the 1.5-MET threshold in some individuals [10].

Measuring Sedentary Behavior

Previous studies in the field of sedentary behavior research used self-reported methods such as questionnaires, which have limited validity and are prone to recall biases and social desirability [17,18]. Furthermore, many studies have used television time as a marker of sedentary behavior to examine adverse health effects [19]. However, television time does not reflect all facets of sedentary behavior, and it is confounded by other factors that are relevant for health outcomes such as dietary intake and socioeconomic status [9]. Therefore, based on advancements in device-based measurements, new paradigms suggest using activity monitors [20]. Currently, an increasing number of studies have used device-based measurements of sedentary behavior [17]. However, the choice of monitor placement is highly important for measuring a sitting or lying position versus a standing posture accurately and, therefore, for meeting the definition stated above. Since the inclinometer measures the angle between the gravity direction and the accelerometer's vertical axis, hip-worn accelerometers are limited to distinguishing between sitting and standing. In contrast, thigh-worn accelerometers are recommended as the gold standard [17,21,22]. Some studies have already used thigh-worn accelerometers: For example, the Maastricht study, which focused on the etiology of type 2 diabetes, its common complications, and its emerging comorbidities, assessed sedentary behavior data from approximately 9000 participants via thigh-worn ActivPALs accelerometers [23]. The Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS) provides a detailed overview of existing studies that have used thigh-worn accelerometers [17]. Although the technical possibilities spur constant progress, this research field is still in its infancy. According to the most recent overview of sedentary behavior and health, there is a pressing need to develop further objective field methods for simultaneously assessing both components of the sedentary behavior definition, which is the postural component (sitting or lying) and the movement intensity and energy component [1].

What is Currently Known About Sedentary Behavior

The latest findings from Stamatakis and colleagues [24] suggest that sedentariness is associated with all-cause and cardiovascular-disease mortality among the least physically active adults. Similar results were found in other epidemiological studies [25,26]. In particular, longer sedentary bouts (ie, a period of uninterrupted sedentary time such as ≥ 30 minutes) may lead to detrimental health effects [27,28].

However, the epidemiological evidence of sedentary behavior' effects on health is incomplete [9]. An important issue is that the majority of studies relied on subjective measures. For instance, a large number of previous studies used self-reported methods such as television time as a marker of sedentary behavior [19,29]. However, Prince and colleagues` [30] meta-analysis revealed that self-reports underestimated sedentary time by 1.74 hours per day in comparison to device-based measures. The evidence for negative health effects relying on device-based measures is sparse as, so far, only very few studies used device-based methods. For example, the EPIC-Norfolk

study has shown that sedentary time was associated with cardiovascular disease, cancer, and all-cause mortality [31]. Accordingly, the evidence regarding the adverse effects of sedentary behavior on health should be interpreted in terms of the problems mentioned above, as different definitions and different measurements naturally lead to different results. While it is indisputable that too much sitting is related to risk factors for health, it remains unclear what "too much" is and what the optimal and practical sedentary break patterns are (ie, type, volume, frequency, intensity, and context) that can buffer negative effects. Therefore, further studies with valid device-based measurements are needed [17,20].

Currently, thigh-worn accelerometers are the method of choice for measuring sedentary behavior accurately [17]. However, accelerometers are unable to provide information about the type or the social and environmental context of sedentary bouts. According to the ecological model [32], sedentary behavior is omnipresent in daily life, and it is multifaceted; for example, it can occur during work, leisure-time, household work, or transport. Moreover, in contrast to physical activity, sedentary behavior is invisible, which means that sedentary behavior is merely a procedural subcomponent of purposeful actions such as working, talking, driving, or reading [33]. To understand everyday sedentary behavior and its antecedents and consequences, it is crucial to collect information about social and environmental contexts. Up to now, we have known little about what everyday sedentary behavior looks like, where, when, and with whom it takes place, and what people are doing while being sedentary. Moreover, to develop effective intervention strategies, it is valuable to know more about socioecological mechanisms within different contexts. Thus, with the aim of changing sedentary behavior patterns, subjective information regarding social and environmental contextual information as well as social-cognitive determinants is a valuable extension for the use of activity monitors [10,18,22].

To the best of our knowledge, there is a lack of studies addressing the social and environmental contexts of sedentary behavior. Fortunately, with EMA, there exists an established approach to assess social and environmental context information in daily life [34]. For example, Liao and colleagues [35] used an EMA design to examine where and with whom children's sedentary behavior occurred during non-school time. Their results revealed that children engaged more in leisure-orientated behavior (such as watching television) than productive sedentary behavior (such as homework). Furthermore, most sedentary time occurred at home and in the company with family members. A further EMA study by Romanzini and colleagues [36] examined sedentary behavior contexts among young adults and showed that the context with the highest occurrence of sedentary behavior was the home context—the main activity while being sedentary was "watching TV and movies," and the main social context was "having alone time." Such pieces of information may enable researchers to tailor context-specific intervention strategies. However, to assess social and environmental contextual information during sedentary episodes or to know when a meaningful moment to intervene occurs, it is crucial to assess variables or to intervene during predefined sedentary episodes (eg, > 20 minutes) and not during other everyday life episodes, in which the person is physically active, for instance. The umbrella term "just-in-time adaptive interventions" (JITAIs) describes interventions that provide behavioral support that corresponds to a need in real-time when the individual is at risk of engaging in an adverse health behavior such as prolonged sedentariness. In particular, this approach comprises a system that offers "just-in-time" automatic behavioral support without individuals' direct participation [37]. A technical solution for a system that detects, triggers, and collects information about prolonged sedentary behavior is to combine accelerometers and EMA (eg, via applications on smartphones) [38,39]. The sedentary behavior–triggered EMA approach enables researchers to incorporate information from subjective measures (eg, questionnaires) and device-based measures (eg, accelerometers) precisely in those situations where the event (such as prolonged sedentary behavior) occurs.

Objectives

The purpose of this paper is (1) to introduce sedentary behavior–triggeredd EMA as a methodological advancement in the field of sedentary behavior research and (2) to examine the accuracy of sedentary behavior–triggered EMA in 3 different studies among healthy adults. Moreover, we compared the accuracy of sedentary behavior–triggered EMA to simulations of random-trigger designs.

Methods

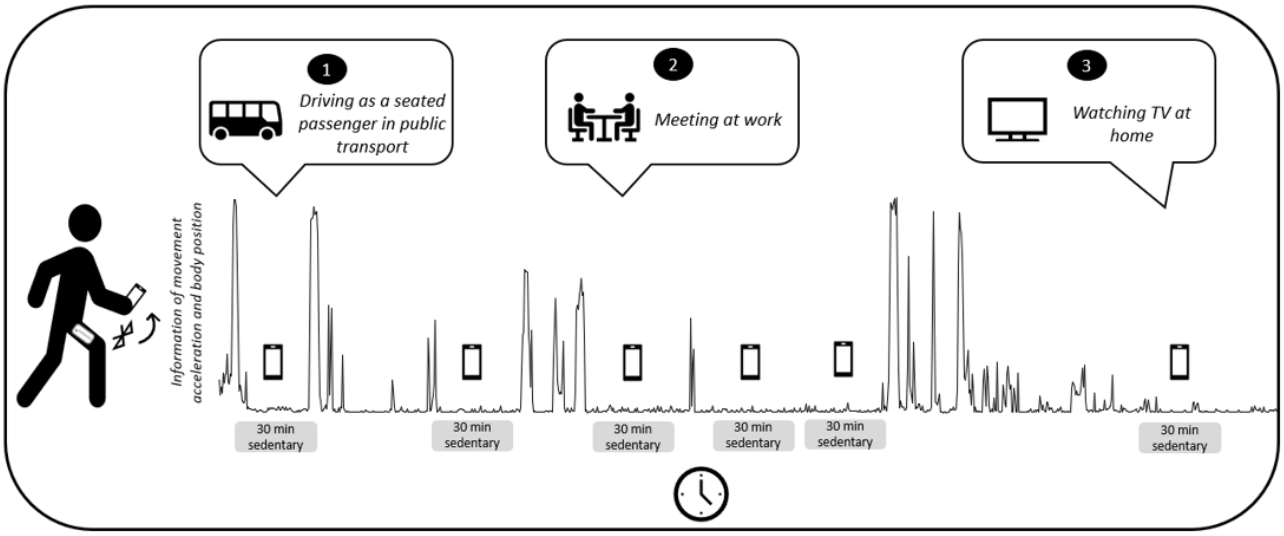
Sedentary Behavior–Triggered Ecological Momentary Assessment

EMA, sometimes also called the Experience Sampling Method (ESM), is currently a state-of-the-art methodology for examining within-subject associations in behavioral relationships [40,41]. Several advantages (such as the ability to assess in everyday life, in real-time, and repeated measurements with a high sampling frequency) have led to the use of EMA in a wide range of research areas [42]. Currently, technological progress enables researchers to collect data in ways that were inconceivable two or three decades ago. For instance, the combination of EMA and external monitors (eg, accelerometers) provides a wide range of new possibilities, such as triggered EMA, a technical evolution within the EMA methodology. This sampling strategy enables researchers to capture specific behavioral episodes such as prolonged sedentary behavior and to ask participants "just in time" about momentary physical and social contexts or psychological parameters such as mood or stress.

The idea of triggered EMA (or e-diaries) is not entirely new, as Ebner-Priemer and colleagues [39] developed a sophisticated activity-triggered algorithm that focused primarily on physically active episodes in everyday life. Based on similar technical requirements, we developed a sedentary-triggered algorithm for which the following equipment is necessary: a thigh-worn accelerometer (eg, Move 3 Activity Sensor, movisens GmbH), an electronic diary (eg, an application on a smartphone), and a technical interface between the e-diary and accelerometer [eg, Bluetooth Low Energy (BLE)] for feedback in real time. In this study, we used the Move 3 accelerometer (movisens GmbH), which is a single-unit accelerometer that captures movement

acceleration and body positions with a measurement range of ±16 g-force (g) at a sampling frequency of 64 Hz. Raw acceleration was stored on an internal memory card. The Move accelerometer has been shown to be a valid device for recording movement behavior [43]. The sedentary behavior–triggered EMA algorithm works as follows: the thigh-worn sensor analyzes data on body position (a sitting or lying position, or an upright position) and transfers the momentary value of the body position in real time to the smartphone. Each time a specific, uninterrupted amount of time spent in a sitting or lying posture is recorded (eg, a 20 or 30 minute period), an e-diary is triggered to begin assessing real-time context information (Figure 1).

Figure 1. Examples of sedentary behavior–triggered Ecological Momentary Assessment (EMA) in everyday life.



Participant Recruitment and Study Design

We used the sedentary behavior–triggered EMA system in 3 different studies, aiming to examine the accuracy of this approach. Table 1 provides an overview of the different study characteristics.

Study 1

We recruited 57 university employees from the Karlsruhe Institute of Technology (KIT) in Germany between May and August 2017. Participants carried a smartphone (Motorola Moto G, Motorola Mobility LLC) and three Move 3 accelerometers for 5 consecutive days. Participants wore accelerometers during the entire measurement period, but not during sleep, swimming, and showering. The thigh-worn monitor and the smartphone were connected via BLE. Sedentary behavior–triggered EMA was used within a mixed sampling scheme. In particular, during the time period from 7:30 am to 9:30 pm, participants received sedentary behavior–triggered prompts (ie, after at least 30 minutes were spent in a sitting or lying position) and randomly triggered prompts at various time points. Since sedentary time is a highly prevalent behavior in daily life, triggered prompts may occur several times per day, which may increase participants' burden. A solution to minimize participants' burden is to implement time-out phases, in which researchers define a time period (eg, of 20, 30, or 40 minutes in duration) when the participants receive no EMA prompts after an answered EMA prompt. During this time-out phase, the study design inhibits EMA prompts, even if the sensor detects an event of uninterrupted sedentary time. In particular, in our first study, EMA prompts occurred no more than every 40 minutes. At each EMA prompt, participants were asked about their social (alone versus not alone) and environmental (home versus work versus leisure activities) contexts (Table 1). Detailed information on the study was described elsewhere [4]. The study was approved by the institutional review board of the Karlsruhe Institute of Technology. All eligible participants received written and oral information regarding the study procedures before written informed consent was obtained. During this procedure, participants were informed of the importance of the smartphone and the accelerometer not losing connection, with a tolerated free-field distance range of approximately 10 meters.

Study 2

We recruited 97 individuals from the University of Konstanz in Germany between May and July 2019. Sedentary behavior was assessed for 4 consecutive days (Thursday to Saturday) using Move 3 accelerometers, which were coupled with smartphones (Motorola Moto G, Motorola Mobility LLC) via BLE. During the time period between 6:00 am and 10:00 pm, short questions were asked via the smartphone whenever the person sat for 20 minutes. We implemented a time-out phase of 20 minutes. At each EMA prompt, participants were asked about their social and environmental contexts (Table 1). The participants completed a short paper-pencil questionnaire before the EMA phase that included demographic variables, age, sex, educational level, height, and weight.

Study 3

We recruited 72 individuals from the University of Konstanz in Germany between January and March 2019. For 4 consecutive days (Monday to Thursday), participants wore a Move 3 accelerometer on their right thigh from the time they got up in the morning to the time they went to bed in the evening. The accelerometer was connected to a smartphone (Motorola Moto G, Motorola Mobility LLC) via BLE. Prior to the assessment,

participants received an extensive briefing on the use of the smartphone and accelerometers and completed a paper-pencil questionnaire that included demographic variables (age, gender, and educational level). During the time period between 6:00 am and 10:00 pm, short questionnaires were asked via the smartphone whenever the person sat for 20 minutes (sedentary trigger). We implemented a time-out phase of 20 minutes. At each EMA prompt, participants were asked about their social and environmental contexts (Table 1).

Data were collected anonymously, and the study fully conformed to the Declaration of Helsinki and the ethics guidelines of the German Psychological Society. Participants received detailed information regarding voluntary participation, the handling of the questionnaires, and the processing of their data, and they gave written informed consent according to the ethics guidelines of the German Psychological Society [44]. According to the guidelines of the ethics committee of the University of Konstanz, the German Research Foundation [45], and the National Science Foundation [46], studies 2 and 3 were exempt from the institutional Ethics Committee review because these 2 surveys were purely observational (noninvasive, noninteractive) and did not induce any type of psychological stress or anxiety, and the participants were not members of a vulnerable group.

Table 1. Study characteristics and Ecological Momentary Assessment (EMA) items.

Study characteristics and EMA items	Study 1 (N=57)	Study 2 (N=97)	Study 3 (N=72)
Characteristics			
Duration (days)	5	4	4
Days of the week	Wednesday-Sunday	Thursday-Sunday	Monday-Thursday
Valid participants[a], n (%)	46 (81)	73 (75)	59 (82)
Gender of valid participants[a], n (%)			
Male	19 (41)	36 (49)	31 (53)
Female	27 (59)	37 (51)	28 (47)
Sedentary trigger (minutes)	30	20	20
Time-out phase (minutes)	Minimum: 40 Maximum: 100	20	20
EMA items			
Environmental context (response options)	*Where are you currently?* (home, work, restaurant, shopping, bus/train, leisure activities, family members, at friends/partners, doctor appointment, exercise, other)	*Where are you currently?* (workplace, canteen, at home, restaurant, bus/train, car, other)	*To which domain would you assign your current sedentary activity?* (work, leisure, home, transport)
Social context (response options)	*Are you alone at the moment?* (yes, no)	*With whom?* (alone, colleagues, friends, family, strangers,other)	*With whom?* (alone, colleagues, friends, family, strangers, other)

[a] ≥ 2 days with ≥ 10 hours wear time.

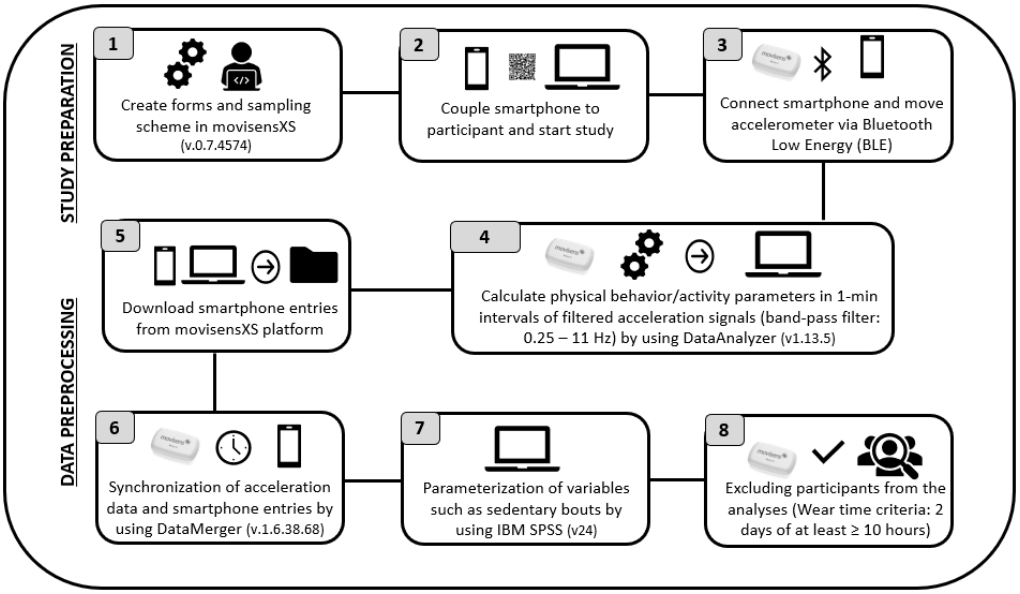
Study Preparation and Data Preprocessing

The same technological system (the Move accelerometer and smartphone with Android operating system) was used in all 3 studies. Thus, from study preparation to data preprocessing, the study procedures (Figure 2) were similar and included the following 8 steps: (1) creation of forms and sampling scheme, (2) coupling of the smartphones to the participants and commencing the study, (3) connection of the smartphones and the accelerometers, (4) processing of the raw acceleration data, (5) downloading the participants' smartphone entries, (6) synchronization of all accelerometer and EMA files into a single data file, (7) parametrization of sedentary-specific variables and calculation of the cumulated sum of the dichotomous variable body position, (8) exclusion of participants who did not fulfill the wear-time criteria (of at least 2 valid days of 10 hours of wear time per day).

First, the sampling scheme and forms (eg, questions about social and environmental context) were created by using the online platform movisensXS (moviesens GmbH). This step included all set-up, such as the selection of the study duration, specification of the trigger option (eg, triggering after 20 minutes or 30 minutes of sitting or lying), and implementation of the time-out triggers. Second, immediately before data collection, the study smartphone was connected to the movisensXS online platform by using the movisensXS app to download the sampling scheme and forms via an individual participant code. Third, the chosen trigger option (eg, triggering after 20 minutes of sitting) was calibrated to the selected body position (the lateral aspect of the right thigh) and connected to the smartphone via BLE by using the movisensXS app. Fourth, after data collection, the recorded raw acceleration data were processed in 1-minute intervals by using the manufacturers' software DataAnalyzer (verson1.13.5, movisens GmbH). During this step, a band-pass filter (0.25-11 Hz) automatically eliminated

gravitational components or artifacts (eg, vibrations when cycling on a rough road surface or sensor shocks). This resulted in an Excel spreadsheet with a self-selected choice of parameters such as body position, movement acceleration intensity (MAI), or activity class. Fifth, the smartphone entries from the participants were downloaded from the movisensXS online platform. Sixth, all accelerometer and EMA files from different participants were synchronized and combined into a single data file using DataMerger (version1.8.0, movisens GmbH). Seventh, prior to the analyses, we parametrized sedentary-specific variables such as sedentary bouts while calculating the cumulated sum of the dichotomous variable body position (1= sitting/lying; 0= upright). Eighth, we excluded participants from the data set if they did not fulfill the wear-time criteria of at least 2 valid days of 10 hours of wear time per day [47]. To distinguish wear-time from nonwear time, we used a commercial algorithm and verified its functioning by scanning simultaneously recorded electrocardiogram parameters [48]. Due to the wear-time criteria, we excluded 11 participants from the sample of study 1, 24 participants from the sample of study 2, and 13 participants from the sample of study 3. More details about the technical system used (the accelerometer and online platform) are described elsewhere [4,49].

Figure 2. Process of study preparation and data preprocessing.



Statistical Analysis

To test the accuracy of sedentary behavior–triggered EMA, we calculated an accuracy score, which is the percentage of all triggered prompts in relation to the total number of all possible triggered prompts. In particular, we first calculated sedentary bouts based on the cumulative sum of the dichotomous variable body position (1= sitting/lying; 0= upright) that was recorded by the accelerometer, and categorized them into the following categories: short bouts (≤ 5 minutes), short-to-moderate bouts (5-19 minutes), moderate-to-long bouts (20-40 minutes), and long bouts (≥ 41 minutes). Second, since earlier studies [43,48] have shown that during sitting or lying periods, the physical activity metric MAI [50] did not exceed the 100 milli g-force (milli-g) threshold, we decided to include only moderate-to-long and long bouts in our analyses if the mean MAI of the bouts was < 100 milli-g. Otherwise, sedentary bouts were categorized without considering acceleration intensity; for instance, a 20-minute bout of cycling in a sitting posture would be incorrectly classified as a sedentary bout. Accordingly, we excluded 9.49% (152/1602) of all moderate-to-long and long bouts in study 1, 4.37% (91/2082) in study 2, and 1.94% (32/1649) in study 3. Third, we calculated the accuracy while checking whether an EMA prompt was triggered during accelerometer-recorded moderate-to-long and long sedentary bouts. We also compared the accuracy of sedentary behavior–triggered EMA with that of a purely random trigger design of (1) every 90 minutes and (2) every 120 minutes. Moreover, we conducted additional analyses to test whether demographic factors influenced the accuracy score and explored the social and environmental context of moderate-to-long sedentary bouts.

Results

Descriptive Statistics

Table 2 presents the descriptive statistics for each study. Across all studies, we analyzed data from 178 participants, 51.4% (91/178) of which were women and 48.6% (87/178) of which were men, with a mean age of 29.25 (SD 10.51; range 19-66) years and an average BMI of 23.23 (SD 3.1; range 17.1-32.4) kg/m^2. Across all studies, participants received a total of 10,771 EMA prompts, which was 60.5 (SD 26.5) EMA prompts per participant. On average, participants answered 54.63 (SD 26.32%; range 4.6-100) of the EMA prompts. According to the accelerometer recordings, participants wore the accelerometer 13.96 (SD 1.41; range 10.2-18.6) hours per day. Of that wear time, participants spent an average of 9.5 (SD 1.74; range 5.49-16.57) hours per day in a sitting or lying position. Our data revealed that 29.3% (5061/17,278) of sedentary bouts were classified as moderate-to-long (2488/17,278, 14.4%) and long bouts (2574/17,278, 14.9%); on average, there were 7.67 (SD

1.91; range 1-13) sedentary bouts of ≥ 20 minutes per participant per day.

Table 2. Participants' characteristics (N=178).

Variable	Study 1[a] (n=46), mean (SD; range)	Study 2[a] (n=73), mean (SD; range)	Study 3[a] (n =59), mean (SD; range)
Age, in years	34.0 (9.6; 25-62)	28.6 (11.6; 19-66)	26.3 (8.5; 21-60)
Gender			
Female, n(%)	27 (59)	37 (51)	28 (48)
Male, n(%)	19 (41)	36 (49)	31 (52)
BMI (kg/m^2)	22.8 (3.3; 17.7-32.1)	23.5 (3.0; 17.1-32.4)	—[b]
Total smartphone prompts[c]	12.31 (1.86; 8-18)	20.72 (7.85; 3-45)	14.83 (5.32; 2-29)
Total triggered prompts[c]	7.3 (2.99; 2-17)	20.72 (7.85; 3-45)	14.83 (5.32; 2-29)
Compliance (%)[d]	79.3 (17.3; 22.2-100)	43.41 (22.5; 4.6-100)	47.49 (23.7; 6.8-93.1)
Wear-time accelerometer (hr/day)[c]	13.6 (1.1; 10.8-16.1)	14.39 (1.6; 10.2-18.6)	13.7 (1.3; 10.7-16.5)
Physical activity of complete measurement period (milli g-force)[c]	86.87 (22.14; 46-148)	81.18 (24.78; 32-141)	79.64 (20.26; 44-155)
Body position: sitting/lying (hr/day)[c]	10.2 (1.6; 7.4-13.7)	9.3 (1.9; 5.7-16.6)	9.16 (1.42; 5.5-12.5)
Total number of short sedentary bouts (≤ 5 min)[c]	11.1 (6.6; 0-29)	10 (6.6; 0-48)	11.9 (6.12; 4-39)
Total number of short-to-moderate bouts (6- ≤ 19 min)[c]	6.3 (2.6; 0-12)	7.1 (2.6; 1-13)	8.6 (3.2; 3-18)
Total number of moderate-to-long bouts (20- ≤ 40 min)[c]	3.5 (1.5; 0-7)	4 (1.8; 0-12)	3.8 (1.2; 2-7)
Total number of long sedentary bouts (≥ 41 min)[c]	4.0 (1.4; 1-7)	4 (1.4; 1-8)	3.6 (1.2; 1-6)

[a]Number of monitoring days per study: Study 1=5 days, Study 2=4 days, Study 3=4 days.

[b]—not available.

[c]Aggregated within the study day per participant.

[d]Percentage of answered Ecological Momentary Assessment prompts across each study sample.

Accuracy

Figure 3 provides a comprehensive overview of the number of accelerometer-recorded sedentary bouts (represented by the black dots on the left side of the figure) as well as a comprehensive overview of the number of bouts that triggered sedentary behavior–triggered EMA (represented by the red dots on the right side of the figure). As a result of a 40-minute time-out trigger in study 1, compared to a 20-minute time-out trigger in studies 2 and 3, there were fewer triggered bouts (red dots) in study 1 than in studies 2 and 3. Moreover, Figure 3 illustrates that the occurrence of sedentary bouts (≥ 20 minutes) is widespread over the day, from morning to evening, in all 3 studies. Overall, 5063 moderate-to-long and long sedentary bouts (≥ 20 minutes) were recorded via accelerometer (Table 3); 11% (559/5057) of these bouts were excluded from the analyses because they occurred prior to or after the study period (ie, 7:30 am - 9:30 pm in study 1, and 6 am - 10 pm in studies 2 and 3). Furthermore, since we implemented a sedentary trigger of ≥ 30 minutes in Study 1, we excluded 32% (464/1450) bouts with a length between 20 and 29 minutes. This resulted in a final number of 4034 sedentary bouts, which could potentially trigger sedentary behavior–triggered EMA. The accuracy calculation revealed that 82.77% (3339/4034) of all possible prompts were triggered. Table 3 summarizes the accuracy on a study level.

Our additional analyses revealed that the sedentary behavior–triggered EMA in the mixed-sampling design of study 1 was 8.97% (n=78) and 20.83% (n=182) higher than that of a simulation of a random-trigger design with prompts every 90 minutes and 120 minutes, respectively. In study 2, the accuracy of the purely sedentary behavior–triggered EMA design was 34.42% (n=587) and 43.46% (n=741) higher than that of a simulation of a random-trigger design with prompts every 90 minutes and 120 minutes, respectively. In study 3, the accuracy of the purely sedentary behavior–triggered EMA design was 34.25% (n=501) and 47.88% (n=699) higher than that of a simulation of a random-trigger design with prompts every 90 minutes and 120 minutes, respectively. These results indicated that the sedentary behavior–triggered EMA system triggered more prompts compared to the simulations of random-trigger designs during moderate-to-long sedentary bouts, and thus, it increases the chance of getting social and environmental context information more often, especially during these kinds of sedentary bouts.

Figure 3. Accuracy of sedentary behavior–triggered Ecological Momentary Assessment (EMA). Left side: the amount of accelerometer-recorded sedentary bouts per study (black dots: sedentary bouts within the study period; grey dots: sedentary bouts outside of the study period). Right side: the amount of triggered EMA diaries (red dots: triggered sedentary bouts; black dots: not-triggered sedentary bouts).

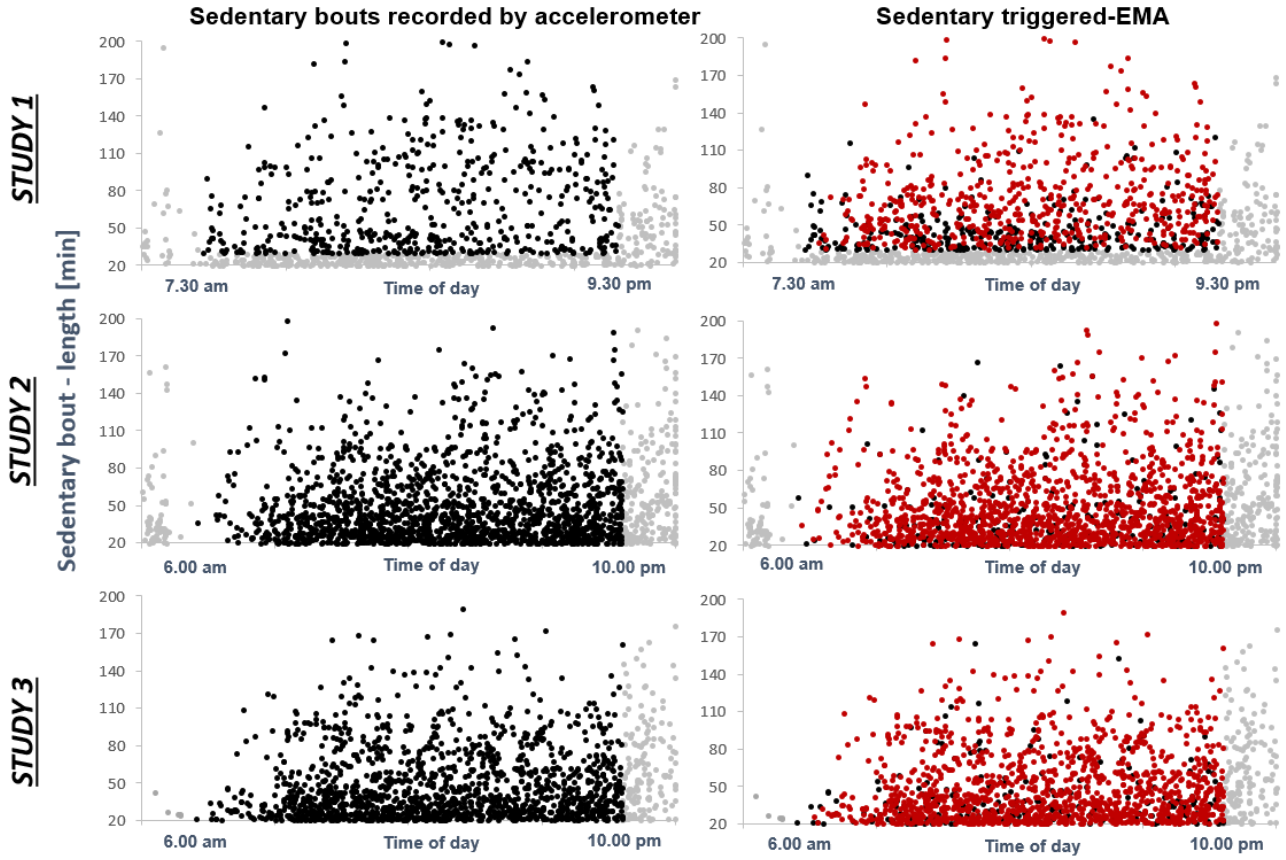


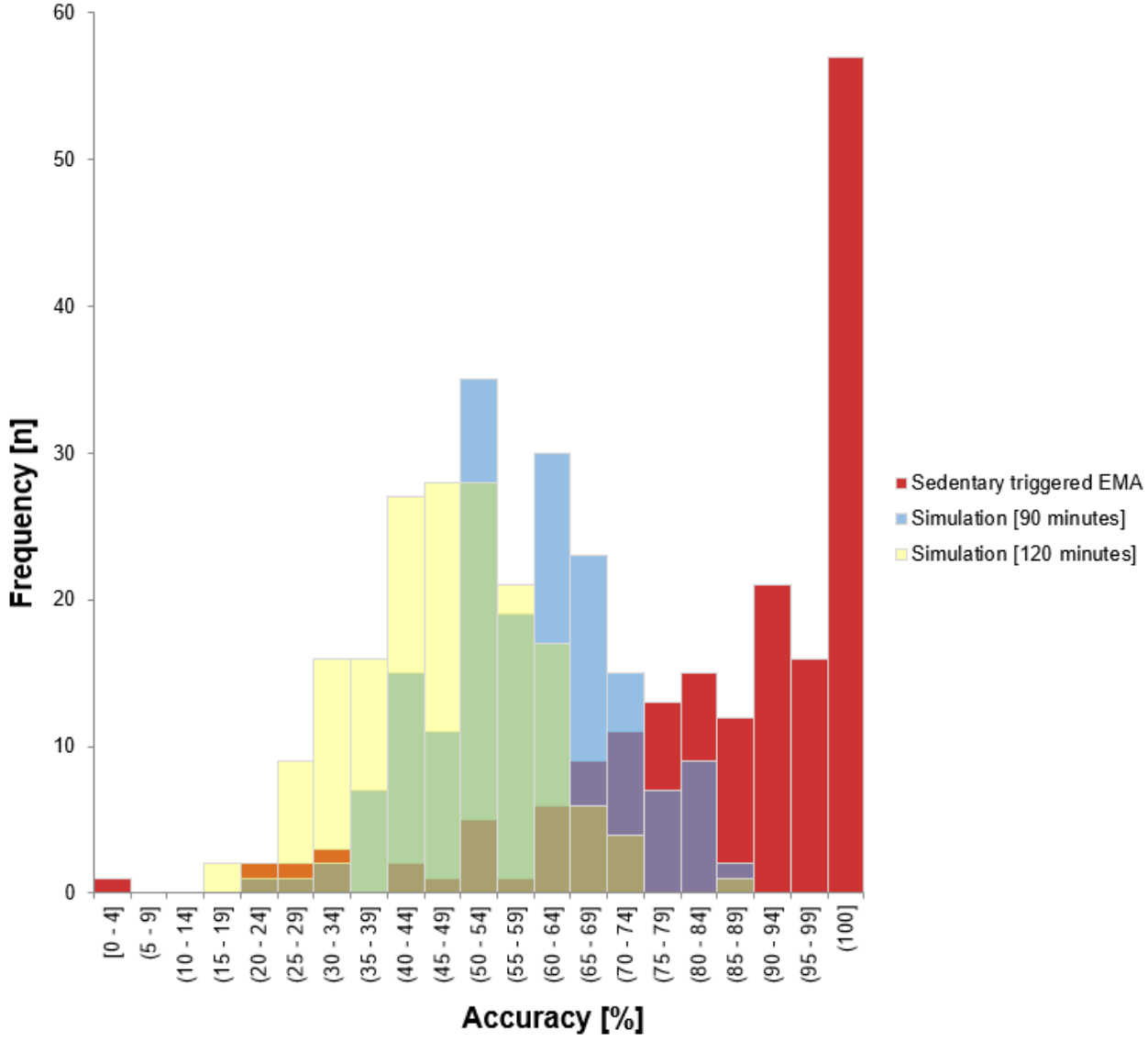
Table 3. Accuracy per study.

Measures	Study 1	Study 2	Study 3
Number of all moderate-to-long sedentary bouts (≥ 20 min) recorded via accelerometer	1450	1993	1614
Sedentary bouts prior to 6 am or 7:30 am	19	70	7
Sedentary bouts after 9:30 pm or 10 pm	98	218	147
Sedentary bouts > 20 - < 30 min	464	N/A[a]	N/A
Total number of bouts that could be triggered	869	1705	1460
Triggered sedentary bouts	617	1434	1288
Accuracy of used study design (%)	71.00	84.11	88.22
Accuracy of 90 min. random triggered simulation (%)	62.03	49.69	53.97
Accuracy of 120 min. random triggered simulation (%)	50.17	40.65	40.34

[a]N/A: Not Applicable

In addition to the accuracy on the study level, we calculated the accuracy per participant. Figure 4 depicts the distribution of the accuracy on the participant level separated by the sedentary behavior–triggered EMA system and simulations of randomly triggered designs of every 90 and 120 minutes. Data analyses revealed a mean accuracy of 80.90 (SD 20.25%; range 0-100%) for the sedentary behavior–triggered EMA system, a mean accuracy of 54.40 (SD 12.56%; range 18.75-83.87%) for the 90-minute simulation, and a mean accuracy of 43.21 (SD 12.02%; range 10.71-81.8%) for the 120-minute simulation. Additional analyses of a 2-tailed *t* test revealed no significant difference (t_{171}= -0.412, *P*=.68) in the accuracy scores for women (80.04, SD 19.59%) and men (81.32, SD 21.37%). Moreover, we found no association between accuracy score and BMI (*r*=0.009; *P*=.92). However, we detected a very small but significant association between accuracy score and age (*r*=-.243; *P*=.001), indicating that the accuracy rate is higher in younger ages.

Figure 4. Distribution of subject-level accuracy separated by sedentary behavior–triggered Ecological Momentary Assessment (EMA) design and simulations of random triggered designs of every 90 and 120 minutes.



Social and Environmental Context

Each time the participants responded to the prompt, they were asked about their current environmental and social context. Across all studies, participants answered 2001 EMA prompts, with an average of 11.57 (SD 7.07) prompts per participant. According to the results about the environmental context, participants reported across all studies that 39.98% (800/2001) of all moderate-to-long and long sedentary bouts occurred during work, 32.93% (659/2001) occurred while at home, 22.44% (449/2001) occurred during leisure activities, and 4.65% (93/2001) occurred during transport. According to the results about the social context, participants reported across all studies that in 56.27% (1126/2001) of all moderate-to-long and long bouts, they were not alone. Specifically, data from studies 2 and 3 revealed that the participants were mostly in the company of friends or family members. Table 4 comprises an overview of the reported results of the environmental and social context by each study on a participant level. The data revealed a high variability between participants. For instance, some participants spent all sedentary bouts during work or while being alone, whereas other participants spent no single sedentary bout during work or while being alone.

In our additional analyses, we found a significant positive correlation (r= 0.4; P<.001) between age and percentage of being with family members during sedentary bouts. Furthermore, we found significant differences (t_{122}=-2.95, P=.004) in the percentage of being with family members for women (24.17, SD 28.46%) and men (11.34, SD 24.17%) during sedentary bouts, indicating that the percentage of moderate-to-long and long sedentary bouts in the company of family members increases with age and is higher for women. However, we found no further significant correlation or differences between age, sex, or BMI and percentage of being in environmental domains (ie, work, home, leisure, or transport) or in other social contexts (ie, colleagues, friends, strangers, or others) during moderate-to-long and long sedentary bouts.

Table 4. Results of social and environmental context for each study sample.

Participant responses	Study 1, mean (SD; range)	Study 2, mean (SD; range)	Study 3, mean (SD; range)
Number of answered prompts	11.02 (5.72; 2-26)	11.87 (7.99; 1-30)	11.78 (6.86; 1-25)
Environmental context (%)[a]			
Work	55.40 (26.64; 0-100)	19.09 (10.43; 0-100)	54.42 (28.85; 0-100)
Home	24.92 (22.42; 0-100)	51.07 (28.23; 0-100)	14.41 (15.55; 0-67)
Leisure	18.79 (20.11; 0-92)	22.73 (22.62; 0-100)	25.99 (22.27; 0-86)
Transport	0.89 (3.03; 0-15)	7.38 (13.78; 0-75)	5.19 (9.82; 0-50)
Social context (%)[a]			
Alone	49.92 (24.90; 0-100)	37.36 (30.61; 0-100)	45.86 (27.73; 0-100)
With colleagues	N/A[a]	12.18 (19.44; 0-83)	17.94 (24.66; 0-100)
With friends	N/A	29.59 (29.06; 0-100)	22.21 (22.26; 0-100)
With family	N/A	23.51 (29.30; 0-100)	11.31 (19.02; 0-100)
With strangers	N/A	1.96 (7.82; 0-60)	6.85 (13.33; 0-56)
With others	N/A	1.13 (4.80; 0-33)	2.48 (6.95; 0-33)

[a]Frequency percentage on a participant level for each study.

[b]N/A: not applicable.

Discussion

Principal Findings

This paper introduced sedentary behavior–triggered EMA as an innovative methodological advancement in the field of sedentary behavior research and assessed the accuracy of sedentary behavior–triggered EMA in 3 different studies of healthy adults. The results indicated that sedentary behavior–triggered EMA captured 82.77% (3339/4034) of all possible sedentary bouts from the different studies. Compared to simulations of random triggered prompts, our data revealed that the sedentary behavior–triggered EMA system triggered more prompts during moderate-to-long sedentary bouts, and thus it increases the chance for getting social and environmental context information more often, especially during these kinds of sedentary bouts. Overall, the results indicate that sedentary behavior–triggered EMA is an accurate method and allows the capture of "just-in-time" social and environmental context information of sedentary behavior bouts.

Enhancing Understanding of Daily Sedentary Behavior

Sedentary behavior has received much attention in the scientific community over the past decade. However, in-depth knowledge about this invisible behavior is still missing [9,33]. Since there is a growing number of studies that have found adverse health effects due to sedentary behavior [1], there is now an urgent need to understand more about circumstances surrounding sedentary behavior such as where it occurs, when it occurs, with whom it occurs, and what people are doing while being sedentary. Thus, high-quality assessment methods such as device-based measurements and methods that collect information on domains (eg, work or leisure), types (eg, watching television while sitting), and contexts (eg, being alone or in company) of behavior are recommended by researchers [10,22]. Only a few studies differentiated among context-specific sedentary times, such as Dempsey and colleagues [51], which have shown that higher sitting time was associated with higher levels of individual biomarkers during television viewing and computer use, and lower levels during occupational sitting. In summary, those few studies mainly differentiated between working and nonworking hours [52,53], whereas the social context remained unconsidered. The social context might be relevant as, for example, the social withdrawal hypothesis [54] reported that greater use of the internet (which is mostly related to a sedentary position) was associated with declines in individuals' social interaction and an increase in depression and loneliness. To verify such a hypothesis, sedentary behavior–triggered EMA may be a useful approach for examining both social interaction and mood in real-time during sedentary bouts (eg, internet use).

In general, EMA is an established procedure for the assessment of intrapersonal and social and environmental contextual information, and it has been widely used in previous studies, for example, in the field of physical activity research [55-58]. To the best of our knowledge, there are very few studies that have applied an EMA design in the context of sedentary behavior research [35,36,59-61]. However, these studies used a random time-based, and not a trigger-based, design. Using only a random time-based design may lead to many prompts being issued during situations other than sedentary bouts. At an extreme, not using sedentary behavior–triggered EMA may impede researchers from unraveling existing associations between sedentary bouts and intrapersonal, interpersonal, and environmental variables (such as mood, social interaction, and context) if, by chance, these variables were assessed only during short sedentary bouts or episodes of physical activity but not during prolonged sedentary bouts. Moreover, since our data revealed a high variability of contextual patterns between participants, the sedentary behavior–triggered EMA design increases the chance of capturing situations that are more

specific. In other words, not using a sedentary behavior–triggered design increases the risk of an incomplete picture of sedentary behavior since we might miss the more rare events (such as sedentary bouts in public transport). Furthermore, in comparison with random triggered designs, sedentary behavior–triggered EMA minimizes the variance of the bout length, which might be helpful to get more contextual information about specific bout lengths and to examine the health effects of different bout lengths (eg, 10, 20, 30, or 60 minutes).

These are the first studies that used a sedentary behavior–triggered EMA and that assessed social and environmental contextual factors during prolonged sedentary bouts. Sedentary behavior–triggered EMA enables researchers to gather relevant information related to the behavior in real-time. Moreover, sedentary behavior–triggered EMA can also be used to unravel dynamic associations. In particular, future researchers may be interested in discovering dynamic associations between sedentary behavior and possible antecedents and consequences, such as the association between sedentary behavior and time-varying constructs like mood, stress, or working memory. In such a study, it may be reasonable to combine triggered and random prompts to maximize the outcome variance. Furthermore, sedentary behavior–triggered EMA can be modified as a methodological system in a JITAI [37]. For example, each time an individual exceeds a specific threshold of time spent in sedentary behavior (eg, ≥ 30 minutes), mobile apps may deliver behavioral support or encouragement to breakup sedentariness, such as by encouraging an individual to stand up and walk for a few minutes. Finally, a triggered EMA study design minimizes not only retrospective bias but also the burden of participants. In particular, participants would be assessed only in situations in which a behavior of interest occurred (eg, prolonged sedentary behavior).

Challenges While Using Sedentary Behavior–Triggered EMA

There are also some challenges when using sedentary behavior–triggered EMA. The accuracy depends on both technical stability and user compliance when participating. In particular, technical issues (such as the accelerometer stopping data recording, or the accelerometer and the smartphone losing their BLE connection or not reconnecting with each other) may hinder a functional system. Furthermore, the compliance and reliability of the participant with regards to carrying the smartphone throughout the study period is a critical aspect. For example, if the participant leaves the smartphone at home while he is going to work, the BLE connection would not be available, and the trigger system would not work. This may explain why the accuracy for some participants was very low in our studies. However, short-term disconnections might be a minor issue for future studies since the next generation of accelerometers can store temporary, online, calculated data and transfers that data to the smartphone after a reconnection. Another issue is that if the participant does not wear the accelerometer and puts the sensor on its side (for example, when in a sitting or lying position), this may lead to the incorrect detection of a prolonged sitting bout. A similar problem may occur if the participant did not wear the accelerometer according to the manufacturer's instructions. However, this could be corrected with valid nonwear time algorithms during offline calculations [48]. Moreover, the study design highly influences the accuracy. Using a longer time-out phase, such as in study 1 (40 minutes), led to a reduced accuracy compared to a shorter time-out phase, such as in studies 2 and 3 (20 minutes). In contrast, the compliance of answered EMA prompts was notably higher in study 1 than in studies 2 and 3. Thus, in summary, it is a fine line between collecting as much data as possible and not burdening a participant to the point of decreasing compliance [62]. This is especially true when the outcome of interest is highly prevalent, as is prolonged sedentary behavior [63]. Therefore, depending on the research question, it could be reasonable to incorporate longer time-out phases. Alternatively, to achieve a high level of adherence, researchers may tailor the sampling scheme by reducing the number of items or the number of study days [64]. Finally, sedentary behavior–triggered EMA increases the chance of getting more contextual information (ie, number of prompts, especially during sedentary bouts) but is still dependent on the compliance of the participants. However, it is possible to combine sedentary behavior–triggered EMA with GPS trajectories [56,57] or wearable camera systems [65] to gather more contextual information.

Conclusions

The results of 3 independent studies revealed that sedentary behavior–triggered EMA is an accurate method for collecting contextual information in daily life. The accuracy of this approach can vary as a function of the study design (eg, time-out triggers), technical stability (eg, connection between the smartphone and accelerometer), and compliance of the participants (eg, following study instructions). Given the growing interest in sedentary behavior research and the lack of knowledge about social and environmental circumstances surrounding sedentary behavior, this sophisticated approach can offer real advancement. Sedentary behavior–triggered EMA can be used to collect social and environmental contextual information or to unravel dynamic associations. Furthermore, it can be modified to develop sedentary behavior–triggered mHealth interventions.

Conflicts of Interest

UEP receives consultancy fees from Boehringer-Ingelheim. MK, CN, and MG have no conflicts of interest to declare.

References

1. Katzmarzyk PT, Powell KE, Jakicic JM, Troiano RP, Piercy K, Tennant B, 2018 Physical Activity Guidelines Advisory Committee. Sedentary Behavior and Health: Update from the 2018 Physical Activity Guidelines Advisory Committee.

Med Sci Sports Exerc 2019 Jun;51(6):1227-1241 [FREE Full text] [doi: 10.1249/MSS.0000000000001935] [Medline: 31095080]
2. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev 2010 Jul;38(3):105-113 [FREE Full text] [doi: 10.1097/JES.0b013e3181e373a2] [Medline: 20577058]
3. Faulkner G, Biddle SJ. Standing on top of the world: Is sedentary behaviour associated with mental health? Mental Health and Physical Activity 2013 Mar;6(1):1-2. [doi: 10.1016/j.mhpa.2013.02.003] [Medline: 32595044]
4. Giurgiu M, Koch ED, Ottenbacher J, Plotnikoff RC, Ebner‐Priemer UW, Reichert M. Sedentary behavior in everyday life relates negatively to mood: An ambulatory assessment study. Scand J Med Sci Sports 2019 May 23;29(9):1340-1351. [doi: 10.1111/sms.13448] [Medline: 31044456]
5. Bailey DP, Hewson DJ, Champion RB, Sayegh SM. Sitting Time and Risk of Cardiovascular Disease and Diabetes: A Systematic Review and Meta-Analysis. Am J Prev Med 2019 Sep;57(3):408-416. [doi: 10.1016/j.amepre.2019.04.015] [Medline: 31377090]
6. Guo C, Zhou Q, Zhang D, Qin P, Li Q, Tian G, et al. Association of total sedentary behaviour and television viewing with risk of overweight/obesity, type 2 diabetes and hypertension: A dose-response meta-analysis. Diabetes Obes Metab 2020 Jan;22(1):79-90. [doi: 10.1111/dom.13867] [Medline: 31468597]
7. US Department of Health and Human Services. Physical activity guidelines advisory committee scientific report. Washington, DC: US Department of Health and Human Services; 2018:2-33.
8. McIntyre L, Dutton M. Eating and activity guidelines for New Zealand adults. Ministry of Health. 2015. URL: https://www.health.govt.nz/system/files/documents/publications/eating-activity-guidelines-for-new-zealand-adults-oct15_0.pdf [accessed 2020-06-01]
9. Stamatakis E, Ekelund U, Ding D, Hamer M, Bauman AE, Lee IM. Is the time right for quantitative public health guidelines on sitting? A narrative review of sedentary behaviour research paradigms and findings. Br J Sports Med 2019 Mar;53(6):377-382 [FREE Full text] [doi: 10.1136/bjsports-2018-099131] [Medline: 29891615]
10. van der Ploeg HP, Hillsdon M. Is sedentary behaviour just physical inactivity by another name? Int J Behav Nutr Phys Act 2017 Oct 23;14(1):142 [FREE Full text] [doi: 10.1186/s12966-017-0601-0] [Medline: 29058587]
11. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
12. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N. Too Little Exercise and Too Much Sitting: Inactivity Physiology and the Need for New Recommendations on Sedentary Behavior. Curr Cardiovasc Risk Rep 2008 Jul;2(4):292-298 [FREE Full text] [doi: 10.1007/s12170-008-0054-8] [Medline: 22905272]
13. Magnon V, Dutheil F, Auxiette C. Sedentariness: A Need for a Definition. Front Public Health 2018 Dec;6:372 [FREE Full text] [doi: 10.3389/fpubh.2018.00372] [Medline: 30622939]
14. Amaro-Gahete FJ, Sanchez-Delgado G, Alcantara JMA, Martinez-Tellez B, Acosta FM, Merchan-Ramirez E, et al. Correction: Energy expenditure differences across lying, sitting, and standing positions in young healthy adults. PLoS One 2019 Jun;14(7):e0219372 [FREE Full text] [doi: 10.1371/journal.pone.0219372] [Medline: 31260490]
15. Henson J, Davies MJ, Bodicoat DH, Edwardson CL, Gill JMR, Stensel DJ, et al. Breaking Up Prolonged Sitting With Standing or Walking Attenuates the Postprandial Metabolic Response in Postmenopausal Women: A Randomized Acute Study. Diabetes Care 2016 Jan;39(1):130-138. [doi: 10.2337/dc15-1240] [Medline: 26628415]
16. Thorp AA, Kingwell BA, Owen N, Dunstan DW. Breaking up workplace sitting time with intermittent standing bouts improves fatigue and musculoskeletal discomfort in overweight/obese office workers. Occup Environ Med 2014 Dec;71(11):765-771. [doi: 10.1136/oemed-2014-102348] [Medline: 25168375]
17. Stamatakis E, Koster A, Hamer M, Rangul V, Lee IM, Bauman AE. Emerging collaborative research platforms for the next generation of physical activity, sleep and exercise medicine guidelines: the prospective physical activity, sitting, and sleep consortium (ProPASS). Br J Sports Med; 2019.
18. Gibbs BB, Hergenroeder AL, Katzmarzyk PT, Lee IM, Jakicic JM. Definition, measurement, and health risks associated with sedentary behavior. Med Sci Sports Exerc 2015 Jul;47(6):1295-1300 [FREE Full text] [doi: 10.1249/MSS.0000000000000517] [Medline: 25222816]
19. Grøntved A, Hu FB. Television viewing and risk of type 2 diabetes, cardiovascular disease, and all-cause mortality: a meta-analysis. JAMA 2011 Jul 15;305(23):2448-2455 [FREE Full text] [doi: 10.1001/jama.2011.812] [Medline: 21673296]
20. Rosenberger ME, Fulton JE, Buman MP, Troiano RP, Grandner MA, Buchner DM, et al. The 24-Hour Activity Cycle: A New Paradigm for Physical Activity. Med Sci Sports Exerc 2019 Mar;51(3):454-464 [FREE Full text] [doi: 10.1249/MSS.0000000000001811] [Medline: 30339658]
21. Janssen X, Cliff DP. Issues Related to Measuring and Interpreting Objectively Measured Sedentary Behavior Data. Measurement in physical education and exercise science ? 2015;19(3):124. [doi: 10.1080/1091367X.2015.1045908]
22. Kang M, Rowe DA. Issues and Challenges in Sedentary Behavior Measurement. Measurement in physical education and exercise science ? 2015;19(3):115. [doi: 10.1080/1091367X.2015.1055566]

23. Schram MT, Sep SJS, van der Kallen CJ, Dagnelie PC, Koster A, Schaper N, et al. The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. Eur J Epidemiol 2014 Jul;29(6):439-451. [doi: 10.1007/s10654-014-9889-0] [Medline: 24756374]
24. Stamatakis E, Gale J, Bauman A, Ekelund U, Hamer M, Ding D. Sitting Time, Physical Activity, and Risk of Mortality in Adults. J Am Coll Cardiol 2019 Apr 30;73(16):2062-2072 [FREE Full text] [doi: 10.1016/j.jacc.2019.02.031] [Medline: 31023430]
25. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc 2009 May;41(5):998-1005. [doi: 10.1249/MSS.0b013e3181930355] [Medline: 19346988]
26. Diaz KM, Howard VJ, Hutto B, Colabianchi N, Vena JE, Safford MM, et al. Patterns of Sedentary Behavior and Mortality in U.S. Middle-Aged and Older Adults: A National Cohort Study. Ann Intern Med 2017 Oct 03;167(7):465-475 [FREE Full text] [doi: 10.7326/M17-0212] [Medline: 28892811]
27. Dempsey PC, Dunstan DW, Larsen RN, Lambert GW, Kingwell BA, Owen N. Prolonged uninterrupted sitting increases fatigue in type 2 diabetes. Diabetes Res Clin Pract 2018 Jan;135:128-133. [doi: 10.1016/j.diabres.2017.11.001] [Medline: 29129482]
28. Dunstan DW, Thorp AA, Healy GN. Prolonged sitting: is it a distinct coronary heart disease risk factor? Curr Opin Cardiol 2011 Oct;26(5):412-419. [doi: 10.1097/HCO.0b013e3283496605] [Medline: 21785350]
29. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. The Lancet 2016 Sep;388(10051):1302-1310. [doi: 10.1016/S0140-6736(16)30370-1]
30. Prince SA, Cardilli L, Reed JL, Saunders TJ, Kite C, Douillette K, et al. A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020 Mar 04;17(1):31-17 [FREE Full text] [doi: 10.1186/s12966-020-00938-3] [Medline: 32131845]
31. Dempsey PC, Strain T, Khaw KT, Wareham NJ, Brage S, Wijndaele K. Prospective Associations of Accelerometer-Measured Physical Activity and Sedentary Time With Incident Cardiovascular Disease, Cancer, and All-Cause Mortality. Circulation 2020 ;141(13):1113-1115.
32. Owen N, Sugiyama T, Eakin EE, Gardiner PA, Tremblay MS, Sallis JF. Adults' Sedentary Behavior. American Journal of Preventive Medicine 2011 Aug;41(2):189-196. [doi: 10.1016/j.amepre.2011.05.013]
33. Gardner B, Flint S, Rebar AL, Dewitt S, Quail SK, Whall H, et al. Is sitting invisible? Exploring how people mentally represent sitting. Int J Behav Nutr Phys Act 2019 Oct 12;16(1):85 [FREE Full text] [doi: 10.1186/s12966-019-0851-0] [Medline: 31606040]
34. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
35. Liao Y, Intille S, Wolch J, Pentz MA, Dunton GF. Understanding the physical and social contexts of children's nonschool sedentary behavior: an ecological momentary assessment study. J Phys Act Health 2014 Mar;11(3):588-595. [doi: 10.1123/jpah.2011-0363] [Medline: 23493261]
36. Romanzini CLP, Romanzini M, Barbosa CCL, Batista MB, Shigaki GB, Ronque ERV. Characterization and Agreement Between Application of Mobile Ecological Momentary Assessment (mEMA) and Accelerometry in the Identification of Prevalence of Sedentary Behavior (SB) in Young Adults. Front Psychol 2019 Apr;10:720 [FREE Full text] [doi: 10.3389/fpsyg.2019.00720] [Medline: 31024382]
37. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
38. Bussmann JB, Ebner-Priemer UW, Fahrenberg J. Ambulatory Activity Monitoring. European Psychologist 2009 Jan;14(2):142-152. [doi: 10.1027/1016-9040.14.2.142]
39. Ebner-Priemer UW, Koudela S, Mutz G, Kanning M. Interactive multimodal ambulatory monitoring to investigate the association between physical activity and affect. Front. Psychol 2013;3:596.
40. Ebner-Priemer UW, Trull TJ. Ambulatory Assessment. European Psychologist 2009 Jan;14(2):109-119. [doi: 10.1027/1016-9040.14.2.109]
41. Kanning MK, Ebner-Priemer UW, Schlicht WM. How to Investigate Within-Subject Associations between Physical Activity and Momentary Affective States in Everyday Life: A Position Statement Based on a Literature Overview. Front. Psychol 2013;4. [doi: 10.3389/fpsyg.2013.00187] [Medline: 23641221]
42. Bolger N, Laurenceau J. Intensive longitudinal methods. New York: Guilford; 2013.
43. Anastasopoulou P, Tubic M, Schmidt S, Neumann R, Woll A, Härtel S. Validation and Comparison of Two Methods to Assess Human Energy Expenditure during Free-Living Activities. PLoS ONE 2014 Feb 28;9(2):e90606. [doi: 10.1371/journal.pone.0090606]
44. Deutsche Gesellschaft für Psychologie e.V. GFPE. Berufsethische Richtlinien. Berlin: Föderation Deutscher Psychologenvereinigungen GbR; 2016.
45. Deutsche Forschungsgesellschaft.. 2019. Ethikvotum. URL: https://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/index.html [accessed 2019-09-19]

46. National SF. Interpreting the common rule for the protection of human subjects for behavioral and social science research. URL: https://www.nsf.gov/bfa/dias/policy/hsfaqs.jsp#exempt [accessed 2019-09-19]
47. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical Activity in the United States Measured by Accelerometer. Medicine & Science in Sports & Exercise 2008;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
48. Barouni A, Ottenbacher J, Schneider J, Feige B, Riemann D, Herlan A, et al. Ambulatory sleep scoring using accelerometers—distinguishing between nonwear and sleep/wake states. PeerJ 2020;8. [doi: 10.7717/peerj.8284]
49. von Haaren B, Ottenbacher J, Muenz J, Neumann R, Boes K, Ebner-Priemer U. Does a 20-week aerobic exercise training programme increase our capabilities to buffer real-life stressors? A randomized, controlled trial using ambulatory assessment. Eur J Appl Physiol 2016 Mar;116(2):383-394. [doi: 10.1007/s00421-015-3284-8] [Medline: 26582310]
50. Van Someren EJ, Lazeron RH, Vonk BF, Mirmiran M, Swaab DF. Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects. J Neurosci Methods 1996 Mar;65(1):55-62. [doi: 10.1016/0165-0270(95)00146-8] [Medline: 8815309]
51. Dempsey PC, Hadgraft NT, Winkler EAH, Clark BK, Buman MP, Gardiner PA, et al. Associations of context-specific sitting time with markers of cardiometabolic risk in Australian adults. Int J Behav Nutr Phys Act 2018 Nov 20;15(1):114 [FREE Full text] [doi: 10.1186/s12966-018-0748-3] [Medline: 30458790]
52. Clemes SA, O'Connell SE, Edwardson CL. Office workers' objectively measured sedentary behavior and physical activity during and outside working hours. J Occup Environ Med 2014 Mar;56(3):298-303. [doi: 10.1097/JOM.0000000000000101] [Medline: 24603203]
53. Thorp AA, Healy GN, Winkler E, Clark BK, Gardiner PA, Owen N, et al. Prolonged sedentary time and physical activity in workplace and non-work contexts: a cross-sectional study of office, customer service and call centre employees. Int J Behav Nutr Phys Act 2012;9(1):128. [doi: 10.1186/1479-5868-9-128]
54. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukophadhyay T, & S. Internet paradox: a social technology that reduces social involvement and psychological well-being? American psychologist ? 1998;53(9):31.
55. Liao Y, Intille SS, Dunton GF. Using Ecological Momentary Assessment to Understand Where and With Whom Adults' Physical and Sedentary Activity Occur. Int.J. Behav. Med 2014 Mar 18;22(1):51-61. [doi: 10.1007/s12529-014-9400-z] [Medline: 24639067]
56. Reichert M, Tost H, Reinhard I, Schlotz W, Zipf A. Salize H-J. Exercise versus nonexercise activitydiaries unravel distinct effects on mood. Med Sci Sports Exerc ? 2017;49(4):73. [doi: 10.1249/MSS.0000000000001149]
57. Koch ED, Tost H, Braun U, Gan G, Giurgiu M, Reinhard I, et al. Mood Dimensions Show Distinct Within-Subject Associations With Non-exercise Activity in Adolescents: An Ambulatory Assessment Study. Front Psychol 2018 Mar;9:268 [FREE Full text] [doi: 10.3389/fpsyg.2018.00268] [Medline: 29563889]
58. Niermann CYN, Herrmann C, von Haaren B, van Kann D, Woll A. Affect and Subsequent Physical Activity: An Ambulatory Assessment Study Examining the Affect-Activity Association in a Real-Life Context. Front Psychol 2016 May;7:677. [doi: 10.3389/fpsyg.2016.00677] [Medline: 27242591]
59. Maher JP, Rebar AL, Dunton GF. Ecological Momentary Assessment Is a Feasible and Valid Methodological Tool to Measure Older Adults' Physical Activity and Sedentary Behavior. Front Psychol 2018 Aug;9:1485 [FREE Full text] [doi: 10.3389/fpsyg.2018.01485] [Medline: 30158891]
60. Kim J, Conroy DE, Smyth JM. Bidirectional associations of momentary affect with physical activity and sedentary behaviors in working adults. Ann Behav Med :2019. [doi: 10.1093/abm/kaz045]
61. Elavsky S, Kishida M, Mogle JA. Concurrent and lagged relations between momentary affect and sedentary behavior in middle-aged women. Menopause 2016 Aug;23(8):919-923 [FREE Full text] [doi: 10.1097/GME.0000000000000645] [Medline: 27326822]
62. 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;104(1):343-351. [doi: 10.1016/S0304-3959(03)00040-X]
63. Hagstromer M, Troiano RP, Sjostrom M, Berrigan D. Levels and Patterns of Objectively Assessed Physical Activity--A Comparison Between Sweden and the United States. American Journal of Epidemiology 2010 Apr 20;171(10):1055-1064. [doi: 10.1093/aje/kwq069]
64. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. Journal of Abnormal Psychology 2020 Jan;129(1):56-63. [doi: 10.1037/abn0000473]
65. Loveday A, Sherar LB, Sanders JP, Sanderson PW, Esliger DW. Technologies that assess the location of physical activity and sedentary behavior: a systematic review. Journal of medical Internet research 2015;17(8):a.

Abbreviations

BLE: Bluetooth Low Energy
EMA: Ecological Momentary Assessment
g: g-force

JITAI: just-in-time adaptive intervention
MAI: movement acceleration intensity
MET: metabolic equivalent
milli-g: milli g-force
SBRN: Sedentary Behavior Research Network

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by Y Liao, J Di; comments to author 10.03.20; revised version received 24.04.20; accepted 03.06.20; published 15.09.20.

Please cite as:
Giurgiu M, Niermann C, Ebner-Priemer U, Kanning M
Accuracy of Sedentary Behavior–Triggered Ecological Momentary Assessment for Collecting Contextual Information: Development and Feasibility Study
JMIR Mhealth Uhealth 2020;8(9):e17852
URL: http://mhealth.jmir.org/2020/9/e17852/
doi:10.2196/17852
PMID:32930668

Original Paper

Effect of Prior Health Knowledge on the Usability of Two Home Medical Devices: Usability Study

Noémie Chaniaud[1], MS; Natacha Métayer[1], PhD; Olga Megalakaki[1], PhD; Emilie Loup-Escande[1], PhD

Centre de Rercherche en Psychologie: Cognition Psychisme et Organisations, Université Picardie Jules Verne, Amiens, France

Corresponding Author:
Noémie Chaniaud, MS
Centre de Rercherche en Psychologie: Cognition Psychisme et Organisations
Université Picardie Jules Verne
Pôle Campus Sud - Bâtiment E - 3ème étage 1, Chemin du Thil - CS 52 501
Amiens, 80 025 CEDEX 1
France
Phone: 33 3 22 82 70 59
Email: noemie.chaniaud@u-picardie.fr

Abstract

Background: Studies on the usability of health care devices are becoming more common, although usability standards are not necessarily specified and followed. Yet, there is little knowledge about the impact of the context of use on the usability outcome. It is specified in the usability standard (ISO 9241-11, 2018) of a device that it may be affected by its context of use and especially by the characteristics of its users. Among these, prior health knowledge (ie, knowledge about human body functioning) is crucial. However, no study has shown that prior health knowledge influences the usability of medical devices.

Objective: Our study aimed to fill this gap by analyzing the relationship between the usability of two home medical devices (soon to be used in the context of ambulatory surgery) and prior health knowledge through an experimental approach.

Methods: For assessing the usability of two home medical devices (blood pressure monitor and pulse oximeter), user tests were conducted among 149 students. A mixed-methods approach (subjective vs objective) using a variety of standard instruments was adopted (direct observation, video analysis, and questionnaires). Participants completed a questionnaire to show the extent of their previous health knowledge and then operated both devices randomly. Efficiency (ie, handling time) and effectiveness (ie, number of handling errors) measures were collected by video analysis. Satisfaction measures were collected by a questionnaire (system usability scale [SUS]). The qualitative observational data were coded using inductive analysis by two independent researchers specialized in cognitive psychology and cognitive ergonomics. Correlational analyses and clusters were performed to test how usability relates to sociodemographic characteristics and prior health knowledge.

Results: The results indicated a lack of usability for both devices. Regarding the blood pressure monitor (137 participants), users made approximately 0.77 errors (SD 1.49), and the mean SUS score was 72.4 (SD 21.07), which is considered "satisfactory." The pulse oximeter (147 participants) appeared easier to use, but participants made more errors (mean 0.99, SD 0.92), and the mean SUS score was 71.52 (SD 17.29), which is considered "satisfactory." The results showed a low negative and significant correlation only between the effectiveness of the two devices and previous knowledge (blood pressure monitor: r=−0.191, P=.03; pulse oximeter: r=−0.263, P=.001). More subtly, we experimentally identified the existence of a threshold level ($\chi^2_{2,146}$=10.9, P=.004) for health knowledge to correctly use the pulse oximeter, but this was missing for the blood pressure monitor.

Conclusions: This study has the following two contributions: (1) a theoretical interest highlighting the importance of user characteristics including prior health knowledge on usability outcomes and (2) an applied interest to provide recommendations to designers and medical staff.

(JMIR Mhealth Uhealth 2020;8(9):e17983) doi:10.2196/17983

KEYWORDS

usability; prior health knowledge; mHealth; home medical devices; blood pressure monitor; pulse oximeter

Introduction

Background

Home medical devices (HMDs) are increasingly being prescribed by health care professionals in order to decongest hospitals, and they are potentially cost-effective approaches of addressing the increasing health care needs [1,2]. HMDs for the general public need to be appropriate for all types of populations regardless of the environment in which they are used [3,4]. Some of them can be very complex to use [5]. However, if HMDs are poorly designed, there could be user errors that could seriously affect patients' safety [3,6-10].

For this reason, a number of ongoing projects are concerned with evaluating and improving the design of these medical devices [10]. User-centered methodologies [11-13] are being increasingly used as new quality and safety standards emerge [14], as a way of avoiding design errors. The "European Conformity" (CE) marking is used to prove safety and usability [15], including *effectiveness*, *efficiency*, and *satisfaction* in the specific context of use of the device. Despite all these standards, usability problems persist. Systematic analyses over the past 10 years have consistently raised an alarm [16-21] by reporting a lack of a framework and a standardized method in usability studies. Although the definition of usability is still an important debate [22], the usability standard has been updated [15] to emphasize that usability is a result of interaction rather than a property of a product [23], which is also defined by its context of use [15], and it includes the following four components: *goals* and *tasks*, *resources*, *environment*, and *users*. These components influence usability results (composed by *effectiveness*, *efficiency*, and *satisfaction*), and it is therefore necessary to know how these components specifically influence usability results. In this study, we focused on *users'* characteristics (by controlling the three other components) according to the metric of usability of the ISO 9241-11 standard [15].

Several researchers have recently investigated the links between users' characteristics and the usability of connected devices in the health field [24-29]. The following four user characteristics seem to be particularly studied in the scientific literature:

- Age: young users outperform older users [3,4,24,25,29-32]
- Experience in information technology (IT) (or technophilia), that is, previous experience in computer and medical devices: technology experts outperform novices [3,24,32,33]
- Motivation: more motivated users outperform less motivated users [34-36]
- Health literacy: users with high levels of health literacy outperform users with low levels of health literacy [29,31,37]

With regard to the relationships between these above-mentioned characteristics and usability, there are only few studies examining the link between health [29,31,37,38] or eHealth literacy [29], particularly prior health knowledge, and usability.

The aim of this study was, consequently, to examine the effects of prior health knowledge on the usability of HMDs. In a within-subjects study design, each participant used a blood pressure monitor and a pulse oximeter.

Related Work

Context of the Study

The study described in this paper is part of the *Smart Angel* project, which aims to provide individual home monitoring for patients who have undergone ambulatory surgery. This monitoring is performed for 1 week, allowing the patient to maintain connectivity with the hospital from home. The device is referred to as an eHealth system. eHealth (or connected health) is defined by Eysenbach [39] as an emerging field bringing together different disciplines such as medical informatics, public health, and business. eHealth offers an important opportunity to address the shortcomings of current health systems and to support health professionals and patients by making them actors in their own health [39,40].

This *Smart Angel* kit is composed of monitoring devices (blood pressure monitor and pulse oximeter) with a digital display application on a touch pad. These two devices were chosen by the medical collaborators of the *Smart Angel* project. They made these choices on the basis of benchmarks by selecting devices with European certification and devices considered easy to use by medical professionals. There is no link between our laboratory and the company manufacturing the equipment. Future patients will find themselves in a postsurgical context (ie, potentially with pain and nausea) at home with the devices and will have to use them (alone or accompanied) three times a day to provide updated health readings to the hospital. User needs data collected from the field in ambulatory surgery have resulted in the selection of connected devices that will monitor patients (blood pressure monitor and pulse oximeter) and that are available on the public market and carry the CE marking.

Usability Assessment

According to ISO 9241-11 [15], usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." This framework makes it possible to stabilize usability around three main dimensions (*effectiveness*, *efficiency*, and *satisfaction*) that are widely used in the field of eHealth [17,21,24,41,42]. Some authors support these standards [43] and the importance of evaluating usability metrics by these three components in independent ways, as well as the collection of both subjective and objective data [17,24,44].

According to the ISO standard [15], effectiveness is defined as the accuracy and completeness with which users achieve specified goals. It is generally measured in terms of the following three points: (1) errors or difficulties in use; (2) unnecessary output elements that interfere with the user's task; and (3) inappropriate decisions made on the basis of inaccurate or incomplete output data. According to the ISO standard [15], efficiency is defined as the resources used in relation to the results achieved (typical resources include time, human effort, cost, and materials). Efficiency includes "the time used" (ie, the time spent trying to achieve an objective).

Satisfaction is defined as the extent to which the user's physical, cognitive, and emotional responses that result from the use of a system, product, or service meet the user's needs and

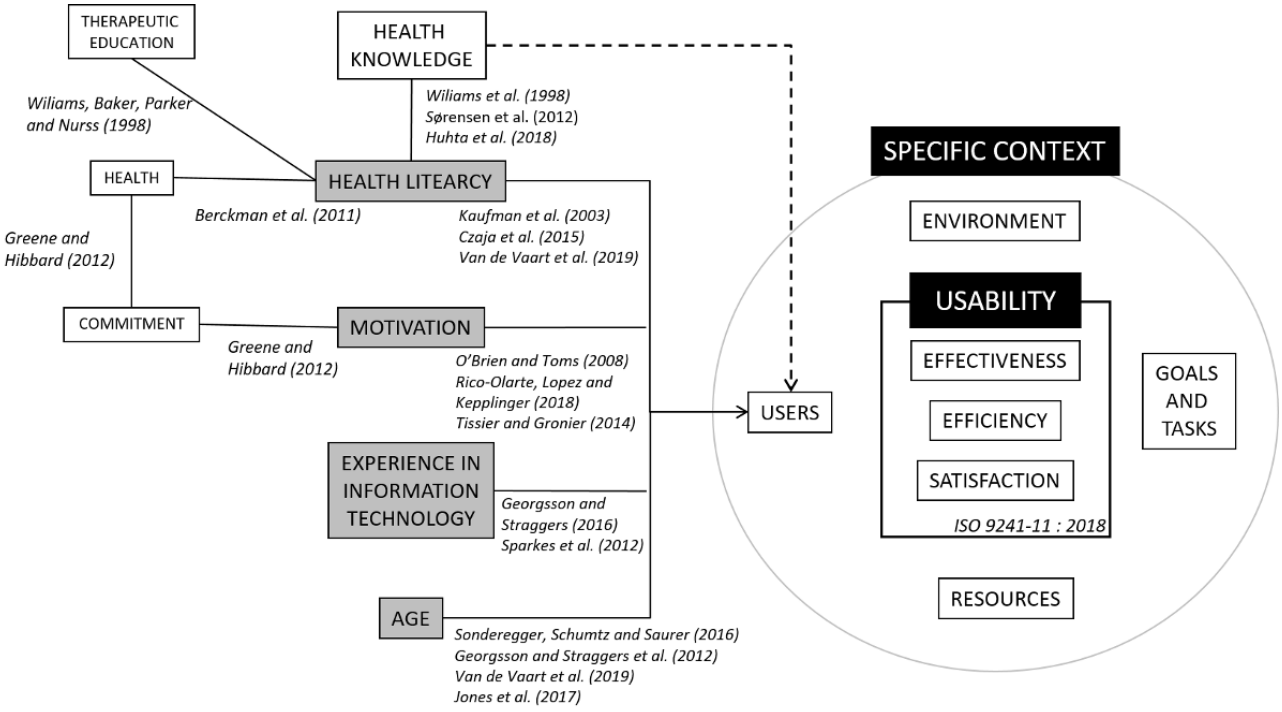
expectations. Satisfaction is assessed by physical (feelings of comfort or discomfort) and cognitive (attitudes, preferences, and perceptions) reactions [15].

Usability and User Characteristics

A better understanding of what influences usability results is of crucial importance to improve the design of medical devices. An increasing number of researchers are taking into account the demographic characteristics of the users they survey in their usability studies, such as age, gender, education level, IT experience, and type of disease [32,45-47]. However, very few collect user skills such as health literacy and device knowledge [29,48]. According to Borsci et al [22] and Grebin et al [49], the lack of attention to human factors is one of the reasons for the slow adoption of medical innovations. These authors proposed to better understand the factors that influence these decision-making processes in order to better understand the resilience abilities of individuals.

We were able to identify, from scientific literature, four main variables of user characteristics that directly influence the usability outcome in a "healthy" population. We propose below a nonexhaustive analysis focusing on user characteristics that influence usability results (Figure 1). These four main variables are age [3,4,24,25,29-32], experience in IT [3,24,32,33], health literacy [29,31,37,38], and motivation [34-36]. For instance, a study by Loorbach et al [50] showed that more motivating manual instructions improved effectiveness and efficiency (but not satisfaction) in relation to a mobile phone in an elderly (age 60-70 years) population. Motivation is directly linked to commitment (activation), which is also directly linked to the patient's health [51]. In addition, age [3,25,29,32] and experience in IT [24,33] impact usability. In contrast, the influences of educational level and professional situations remain ambivalent [24,26,29].

Figure 1. The influences of the four main user characteristics filled in grey (age, experience in information technology, motivation, and health literacy) from the scientific literature on usability results (effectiveness, efficiency, and satisfaction), which have an impact on patients' health. Solid lines are inferred from published literature, and dashed lines are hypothetical. The solid curved line is the ISO 9241-11:2018 metric.



Usability, Health Literacy, and Prior Health Knowledge

Health literacy seems to be an area on which many researchers are focusing to improve the design of medical devices [27,37,38,52] and thus improve the health of patients [42]. Czaja et al [37] showed that people with a low level of health literacy had difficulty completing electronic personal health records. Mackert et al [53] found that patients with low literacy levels use less health IT. Kim and Xie [38] conducted a systematic review of the impact of health literacy on health technologies. After an analysis of 74 studies, the authors concluded that the major barrier to access and use online health information for individuals with low literacy levels is strongly related to usability. A low level of health literacy can limit patients' access to therapeutic education [54,55] and consequently lessen their commitment [56]. Paasche-Orlow and Wolf [57] even showed that literacy is directly related to health outcomes.

Even with a high level of basic literacy, a person may have difficulty obtaining, understanding, and using health information [58]. Health literacy is especially complex to assess, as studies have shown that it is not related to the age, revenue, or education of patients [59], and there is still no consensus on its definition [60]. However, most of the definitions include prior health knowledge in the concept of health literacy [60,61]. Williams et al [55] demonstrated the link between knowledge and health literacy through experimentation with patients having chronic diseases. It would appear that it is necessary to use health knowledge about how the body and treatments work in order to use health information. Thus, it is possible to suggest that

better knowledge of health could lead to better usability of medical equipment.

Knowledge is often defined as a belief that is true and justified. A correct or incorrect answer is interpreted as simply meaning that a person knows or does not know something. It is necessary to integrate the "test taker's certainty" [62] into the test in order to take into account all dimensions of knowledge. We are seeking to specifically measure knowledge about the human body in relation to the medical devices chosen, that is, to understand the functioning of arterial pressure in relation to the use of a blood pressure monitor and to understand the functioning of blood oxygenation in relation to the use of a pulse oximeter.

Aim and Hypotheses

As we mentioned previously, theoretical studies are needed to better understand how users' characteristics influence usability in order to design safer devices for patients. The influences of age, experience, literacy levels, and motivation on usability results have been proven (Figure 1). To our knowledge, no study has highlighted the empirical link between health knowledge and usability results. Thus, we propose to analyze this relationship by controlling the three other variables (ie, we conducted a study in a controlled environment, and we selected a young, healthy, and technologically familiar population) to limit the impact on usability results.

The aim of our experimental study was to examine the relationship between prior health knowledge and usability of HMDs. We hypothesize that participants with good prior health knowledge will use these devices with better *effectiveness* (H1), *efficiency* (H2), and *satisfaction* (H3).

Methods

Recruitment

One hundred and fifty-three psychology undergraduate students (mean age 20.72 years, SD 1.65 years; age range 18-32 years; 41 male and 112 female students) at the University of Picardie Jules Verne in Amiens (France) participated in the experiment. They were recruited in the university hall in March and April 2018 and were informed that they were testing medical equipment. Four participants were removed from the analysis because of technical problems. All participants were native French speakers and signed an informed consent form. The data collected on participants were anonymous. This research complied with the American Psychological Association Code of Ethics. Full review and approval were not required according to our institution's guidelines and national regulations. Participants did not receive any financial compensation for their participation. We chose this population to avoid age bias [25] and to have homogeneous abilities in IT/computer knowledge [24]. In addition, this type of population has very little experience in the use of medical devices, making it possible to stabilize our results and avoid population-related biases.

Materials and Measures

Medical Devices and Tasks

Participants used the following two medical devices intended for use by the general public (Figure 2): a wireless pulse oximeter (iHealth Oximeter PO3) and a wireless wrist monitor measuring blood pressure (iHealth BP7). The blood pressure monitor, once the measurement is complete, displays the systolic (*SYS*) and diastolic pressure (*DIA*) in mmHg, as well as the pulse rate next to a heart-shaped pictogram with the label *PUL*. The pulse oximeter indicates the oxygen level in the blood in $\%SpO_2$, as well as the pulse rate in PR bpm. The measurements and indicators (in French; eg, bpm) of each device remain displayed for a few seconds.

Ten tasks for the use of the blood pressure monitor and nine tasks for the use of the pulse oximeter were involved (Table 1).

Figure 2. The devices used. (A) The wireless blood pressure wrist monitor (iHealth BP7). (B) The pulse oximeter (iHealth Oximeter PO3).

Table 1. User tasks for the blood pressure monitor and pulse oximeter.

Task	Blood pressure monitor	Pulse oximeter
Task 1	Turn on the device	Turn on the device
Task 2a	Position the device correctly	Position the device correctly
Task 2b	Tilt the forearm using a system of illuminated arrows to help find the correct elbow angle to start the measurement	N/A[a]
Task 3	Start the measurement	Start the measurement
Task 4	Remain still during the measurement	Remain still during the measurement
Task 5	Record the measurement on a sheet	Record the measurement on a sheet
Task 6	Interpret the symbols on the device (eg, SYS for systolic, DYA for diastolic, and bpm for beats per minute)	Interpret the symbols on the device (eg, SPO2 for oxygen level and bpm for beats per minute)
Task 7	Interpret the results	Interpret the results
Task 8	Remove the device	Remove the device
Task 9	Turn off the device	Turn off the device

[a]N/A: not applicable.

Questionnaires

Participants were asked to answer the following questionnaires before and after accomplishing the task:

- Health knowledge questionnaire (before the task):
  For prior evaluation, we used a questionnaire designed in collaboration with medical and education professionals (Multimedia Appendix 1). Participants were asked to mark a number of statements as true, false, or I don't know (eg, The heart acts like a pump). We propose a translated version (Multimedia Appendix 2) of the questionnaire in English, which requires validation.
- Personal information/use of technology questionnaire (before the task):
  This questionnaire comprised three personal details (ie, age, gender, and education level) and two items (adapted from Agarwal and Prasad [63]) related to the participant's use of and willingness to explore innovations in IT (eg, Which of these technologies do you use and how often?). On a five-point Likert scale, the possible answers ranged from never to very often.
- System usability scale (SUS) (after the task):
  This 10-item survey aimed at recording subjective assessments of usability [64,65] is a "quick and dirty" tool with five response options from strongly agree to strongly disagree. We used the modified version described previously [66]. This version was further modified by changing the word "system" to "medical device." This type of change has no impact on the validity or reliability of the survey instrument [67].

Procedure

For participants, the session was divided into two parts. Each part took approximately 15 minutes. In the first part, participants completed the *personal information/use of technology* questionnaire followed by the *health knowledge* questionnaire (Multimedia Appendix 1). Participants were allowed to rest for the blood pressure measurement that followed. In the second part, they were invited by experimenters to manipulate two HMDs (oximeter and blood pressure monitor) in a randomized order. The only instruction given to the participants was to perform measurements on themselves and to record the results on their handover sheet. The participants were filmed during the session. The experiment was conducted with one participant per session in a quiet room with minimal distraction. Participants did not receive any previous training or demo from the experimenter. The experimenter only intervened when there were technical problems (eg, a battery problem). At the end of each manipulation, participants answered questions about their measurements, including questions about how to interpret the data, and they completed the SUS questionnaire (see above).

Scoring and Statistical Analysis

The recordings were analyzed by two expert evaluators using BORIS (Behavioral Observation Research Interactive Software) [68]. ISO 9241-11 measures of effectiveness, efficiency, and satisfaction were assessed as presented below.

Objective Measures of Usability

In order to measure effectiveness (ie, number of handling errors) and efficiency (ie, handling time), we used the same metrics as those in the studies by Georgsson and Staggers [24] and Sheehan and Lucero [45].

The measure of effectiveness was analyzed by rating the number of handling errors, such as not putting the monitoring cuff on in the right position. Four main errors were detected for the blood pressure monitor and three for the oximeter (for instance, the participant did not hang the monitor in the right position). A scoring grid was used to identify handling errors. Participants sometimes repeated the same error several times, and we recorded the cumulative number of handling errors. The number of handling errors was between 0 and 4, except when participants repeated the same error successively. All errors were then averaged out.

The measure of efficiency was analyzed by the handling time (in seconds). The handling time was measured from the time participants first touched the device to the time they turned it off after taking the measurement.

Subjective Measures of Usability

In order to measure satisfaction, we used the SUS, as in other studies [24,26,69]. Scores were calculated according to Brooke guidelines [64]. The SUS score ranges from 0 to 100, with lower scores indicating lower usability.

Measuring Prior Health Knowledge

The *health knowledge* questionnaire (Multimedia Appendix 1) included 40 items and was divided into two parts (20 items concerning blood pressure and 20 concerning blood oxygenation). In this questionnaire, each correct answer corresponded to 1 point and each incorrect answer corresponded to 0 points, with a total score between 0 and 40.

Health Data Recorded

We considered whether participants were able to correctly record their blood pressure and the rate of oxygenation and whether they were able to annotate the first two measurements indicating their blood pressure and pulse oximeter. The health data recorded by the HMDs are not explicitly given. For example, participants should read three measurements on the blood pressure monitor (the first two are systolic and diastolic blood pressures and the third is pulse). Each correct measure recorded (eg, blood pressure and oxygenation rate) corresponded to 1 point, and each incorrect measure recorded corresponded to 0 points, with a total score between 0 and 1 for the blood pressure monitor and between 0 and 1 for the pulse oximeter.

Data Analysis

Results were computed using SPSS version 22 (IBM Corporation). Descriptive results were compared with participants' health knowledge. For comparisons among user characteristics, user performance, and satisfaction, we compared gender, age, and IT/computer knowledge and experience (based on rated experience vs rated inexperience) against effectiveness, efficiency, and SUS mean scores for the blood pressure monitor and for the pulse oximeter. The sample did not follow normal distribution; therefore, the correlation coefficient on ranks (Spearman ρ) was used between ISO 9241-11 metrics and health knowledge. We then established three knowledge clusters (low, medium, and high levels of health knowledge) using *k*-means. The Kruskal-Wallis nonparametric test was used to analyze these clusters.

Results

Interjudge Reliability

We used the intraclass correlation coefficient (ICC) to verify interjudge reliability for quantitative data [70,71]. Double coding was performed on 33% of the video data collected by two independent researchers specialized in cognitive psychology and cognitive ergonomics. The average ICC measure for blood pressure monitor errors was 0.962 (95% CI 0.933-0.979; $F_{47,47}$=26.461; P<.001). The average ICC measure for the handling time of the blood pressure monitor was 0.995 (95% CI 0.992-0.997; $F_{47,47}$=261.275; P<.001). The average ICC measure for pulse oximeter errors was 0.936 (95% CI 0.887-0.964; $F_{47,47}$=15.732; P<.001). The average ICC measure for the handling time of the pulse oximeter was 0.995 (95% CI 0.992-0.997; $F_{47,47}$=261.275; P<.001).

Internal Consistency of the Prior Health Questionnaire

The evaluation method commonly used to assess the reliability of a test is the test-retest method. This method consists of administering the same test twice in the same individuals separated in time. If the test is consistent, the correlation between the test and retest will be high. However, this method supports stability in the evaluation and therefore seems inappropriate in the case of our knowledge questionnaire.

Indeed, knowledge can fluctuate through time. Therefore, a method of assessing reliability that does not involve a double test is necessary in the case of our study. The internal consistency was notably estimated by the Cronbach α coefficient. It is considered good when the Cronbach α value is above .70 [72]. The health knowledge questionnaire was pretested on a population of 68 undergraduate students in psychology, which enabled us to confirm a standardized distribution of results (mean 20, median 20 [50% of correct answers], SD 5.3). The Cronbach α values were .72 for the blood pressure knowledge scale and .76 for the blood oxygenation knowledge scale, which can be considered good for both scales.

User Statistics: Participant Demographics, IT Experience, and Medical Device Experience

Among the 149 participants (Table 2), we decided to retain only those who successfully used the medical devices in order not to bias efficiency results (ie, 137 for the blood pressure monitor and 147 for the pulse oximeter). Indeed, if a user handles the device for a very long time before definitively dropping it, it is not possible to take into account these measures of handling time (efficiency) because this will not correspond to the measurement criteria of this variable. We began by analyzing the impacts of user characteristics (eg, demographics, IT experience, and medical device experience) on usability. We then analyzed the correlation between the participant's health knowledge and usability to test our hypotheses (H1, effectiveness; H2, efficiency; and H3, satisfaction) as follows: (1) for the blood pressure monitor, (2) for the pulse oximeter, and (3) the cluster analyses.

Age and level of education had no impact on usability results owing to the similar profiles of students (except gender) for effectiveness (U=1686.5, P=.02) and satisfaction (U=2756, P=.01) in the case of the pulse oximeter. The Kruskal-Wallis test revealed no significant difference between IT experience and usability results (effectiveness: $\chi^2_{2,136}$=0.5, P=.77; efficiency: $\chi^2_{2,136}$=0.6, P=.73; satisfaction: $\chi^2_{2,136}$=0.2, P=.88) for the blood pressure monitor. Additionally, for the pulse oximeter, no significant difference was observed between IT experience and usability results (effectiveness: $\chi^2_{2,146}$=1.8, P=.40; efficiency: $\chi^2_{2,146}$=1.7, P=.43). However, participants who had already used a pulse oximeter were significantly more satisfied because they had a better SUS score (satisfaction: $\chi^2_{2,146}$=6.036; P=.049). Details of these measures are available in Multimedia Appendix 3.

Table 2. Sociodemographic characteristics and experience (information technology and medical device) of the study cohort.

Variable	Value (N=149), n (%) or mean (SD)
Sociodemographic characteristics	
Age (years)	20.72 (1.65)
Gender	
Male	37 (24.8%)
Female	112 (75.2%)
Education level	
First year	2 (1.3%)
Second year	88 (59.1%)
Third year	59 (39.6%)
Information technology experience	
High	59 (43.1%)
Medium	47 (34.3%)
Low	31 (22.6%)
Comfort with computer use	
High	88 (59.0%)
Medium	60 (40.3%)
Low	1 (0.7%)
Frequency of computer use	
Every day (very often)	103 (69.1%)
Several times a week (often)	39 (26.2%)
Once in a while (seldom)	6 (4.0%)
Never	0 (0%)
Comfort with cellphone use	
High	120 (80.5%)
Medium	29 (19.5%)
Low	0 (0%)
Frequency of cell phone use	
Every day (very often)	125 (83.9%)
Several times a week (often)	19 (12.8%)
Once in a while (seldom)	4 (2.7%)
Never	0 (0%)
Connected device comfort	
High	10 (6.7%)
Medium	31 (20.8%)
Low	108 (72.5%)
Frequency of connected device use	
Every day (very often)	3 (2.0%)
Several times a week (often)	4 (2.7%)
Once in a while (seldom)	46 (30.9%)
Never	95 (63.8%)
Experience with the use of medical devices	
Blood pressure monitor	

Variable	Value (N=149), n (%) or mean (SD)
Yes	86 (57.7%)
No	63 (42.3%)
Pulse oximeter	
Yes	20 (13.4%)
No	129 (86.6%)

Evaluation Outcomes

Blood Pressure Monitor

Usability Testing

One hundred and thirty-seven users were able to use the device successfully, that is, to obtain a measure (Table 3). Twelve participants gave up owing to the complexity of using the device. This result indicates a lack of effectiveness. For the rest of the sample, users made a mean of 0.77 errors (SD 1.49). Users manipulated the device for a mean of 260.91 seconds (SD 107.12). The mean SUS score was 72.4 (SD 21.07), which is "satisfactory" [67]. Regarding the descriptive results, this device appeared not to be at all user friendly for a population that has had no previous training. In addition, the data correctly recorded were extremely poor. Only 12 (9%) participants were able to correctly record their blood pressure.

Table 3. Results of the usability measurements of the blood pressure monitor and Spearman correlation between usability and prior health knowledge (N=137).

Measurement	Value		Correlation with health knowledge (*r*)	*P* value (unilateral)
	Mean (SD)	Range (minimum-maximum)		
Effectiveness (errors)	0.77 (1.49)	0-8	−0.191	.03
Efficiency (seconds)	260.91 (107.12)	90.5-681	−0.104	.23
Satisfaction (SUS[a] score)	72.4 (21.07)	27.5-100	0.146	.09
Data recorded (rate of correct response)	0.099 (0.28)	0-1	0.302	<.001

[a]SUS: system usability scale.

Correlation of Usability With User Health Knowledge

Participants scored a mean of 10.16 out of 20 (SD 2.95, range 1-16), with 50.8% (n=69) correct answers. Spearman correlation between blood pressure knowledge and usability metrics showed low and negative but significant correlations between the number of errors and participant knowledge (r=−0.191, P=.03). Participants with a high level of health knowledge made fewer errors. The results were however not significant for efficiency (r=−0.104, P=.22) and satisfaction (r=0.146, P=.08). The majority of participants (n=12) misrecorded their data. We observed that only participants with a high level of health knowledge were able to record their blood pressure correctly (r=.302, P<.001). This is a medium, positive, and significant correlation that illustrates the limitations of this medical device. Although most participants knew how to manipulate the blood pressure monitor, they struggled to read the results.

Pulse Oximeter

Usability Testing

One hundred and forty-seven participants were able to use the pulse oximeter successfully (Table 4). Only two participants gave up because they failed to use the device properly, and their results were excluded from the analysis. The oximeter therefore appears to be easier to use than the blood pressure monitor, which was abandoned by 12 participants, but participants made more errors (mean 0.99, SD 0.92). The mean SUS score was 71.52 (SD 17.29), which is "satisfactory" [67]. In addition, data readings were quite good. On average, 64.6% (n=95) of the participants were able to record their oxygen levels correctly.

Table 4. Results of the usability measurements of the pulse oximeter and Spearman correlation between usability and prior health knowledge (N=147).

Measurement	Value		Correlation with health knowledge (*r*)	*P* value (unilateral)
	Mean (SD)	Range (minimum-maximum)		
Effectiveness (errors)	0.99 (0.92)	0-6	−0.263	.001
Efficiency (seconds)	158.42 (75.75)	24.8-458.8	−0.062	.45
Satisfaction (SUS[a] score)	71.52 (17.29)	15-100	0.195	.02
Data recorded (rate of correct response)	0.65 (0.48)	0-1	0.018	.82

[a]SUS: system usability scale.

Correlation of Usability With User Health Knowledge

For health knowledge of blood oxygenation, participants scored a mean of 5.28 out of 20 (SD 2.77, range 0-15), corresponding to 26.3% correct answers. Spearman correlation between blood pressure knowledge and ISO usability metrics showed similar results for the pulse oximeter and the blood pressure monitor. There was a low, negative, and significant correlation between the number of errors and participant knowledge (r=−0.263, P=.001). We may deduce that participants with better health knowledge make fewer mistakes and are therefore more efficient in handling the device. Significance for satisfaction and participant knowledge was also observed (r=0.195, P=.02). However, there was no correlation when reading the results (r=0.018, P=.82).

Cluster Analysis

In view of the poor correlations, we considered the possibility of a threshold effect. For this purpose, we created clusters (k-means), separating participants into three assignment groups according to their level of prior health knowledge. We then looked at the three knowledge groups (low, medium, and high) in terms of their *effectiveness*, *efficiency*, and *satisfaction* in relation to the blood pressure monitor (Table 5) and the pulse oximeter (Table 6).

Table 5. Prior health knowledge measurements by cluster groups according to usability metrics for the blood pressure monitor using the Kruskal-Wallis test (N=137).

Measurement	Low group (N=26; overall mean 8.73/40, SD 2.51)		Medium group (N=66; overall mean 14.47/40, SD 1.72)		High group (N=45; overall mean 20.93/40, SD 2.9)		Kruskal-Wallis test	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	χ^2 (*2,136*)	*P*
Effectiveness (errors)	92.00 (1.44)	0-7	0.86 (1.62)	0-8	0.56 (1.01)	0-4	2.5	.29
Efficiency (seconds)	286.42 (110.16)	112.1-534.3	257.53 (119.41)	90.5-681.04	251.14 (83.72)	127.5-494.2	2.9	.23
Satisfaction (SUS score)	70.38 (16.46)	27.5-100	72.39 (16.46)	27.5-97.5	73.82 (16.44)	40-100	1.0	.60
Health data read	0.00 (0.00)	0-0	0.05 (0.21)	0-1	0.20 (0.41)	0-1	3.5	.18

Table 6. Prior health knowledge measurements by cluster groups according to usability dimensions for the pulse oximeter using the Kruskal-Wallis test (N=147).

Measurement	Low group (N=44; overall mean 9.77/40, SD 2.76)		Medium group (N=29; overall mean 14.1/40, SD 0.77)		High group (N=74; overall mean 19.31/40, SD 3.28)		Kruskal-Wallis test	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	χ^2 (*2,146*)	*P*
Effectiveness (errors)	1.34 (0.88)	0-4	0.69 (0.6)	0-7	0.92 (1.04)	0-6	10.9	.004
Efficiency (seconds)	168.38 (76.98)	48.2-385.4	163.68 (77.47)	71.4-452.5	150.44 (74.49)	458.8-7	2.0	.37
Satisfaction (SUS score)	69.86 (16.38)	32-95	71.45 (23.23)	15-98	72.73 (15.16)	30-100	7.5	.02
Health data read	0.89 (0.32)	0-1	0.86 (0.35)	0-1	0.91 (0.29)	0-1	0.03	.98

Concerning effectiveness, the Kruskal-Wallis test revealed a significant difference between the number of pulse oximeter handling errors (effectiveness) and cluster groups ($\chi^2_{2,146}$=10.9, P=.004). The low group (mean 1.3, SD 0.88) had significantly more errors than the other two groups. We performed the same analysis with efficiency, taking into account time and satisfaction using SUS measures. Efficiency did not reveal any significant difference ($\chi^2_{2,146}$=2.0, P=.37). However, this

threshold effect was found for satisfaction based on the SUS score ($\chi^2_{2,146}$=7.5, *P*=.02), probably because of the number of errors made for the pulse oximeter. These results were not transferable to the blood pressure monitor.

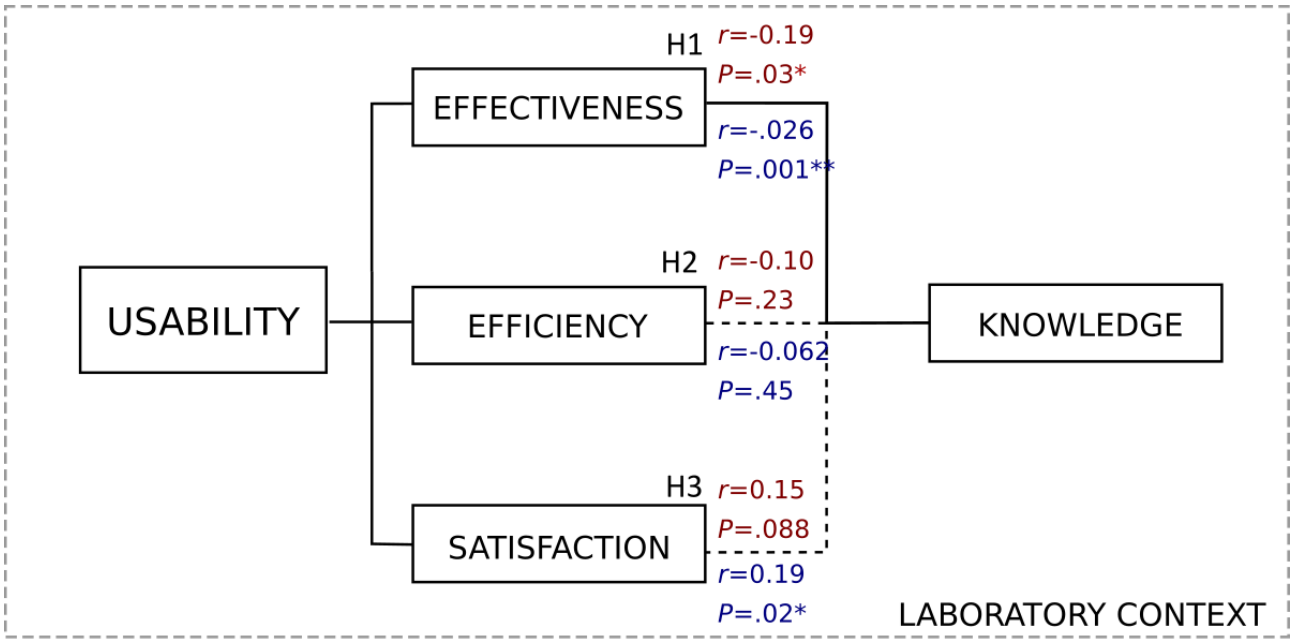
Discussion

Main Contributions

The objective of this study was to explore how prior health knowledge, seen as part of the health literacy level [55,60,73], could impact the usability results (effectiveness, efficiency, and satisfaction) of HMDs. The findings support the central hypothesis of this study, namely that *better health knowledge leads to better usability*. Participants with good knowledge were more effective than those without good knowledge. More precisely, participants with knowledge of how blood pressure works in the human body made significantly (*P*=.03) fewer handling errors when using the blood pressure monitor. Having a basic understanding of how the body works, such as blood pressure, would help to better understand how the blood pressure monitor works and, for instance, prevent posture errors. This was the case for both devices, but a threshold effect was visible in the case of the oximeter. User tests also indicated that the blood pressure monitor was more difficult to use than the oximeter. Reading the result of a measurement seems to be intuitive in the case of the oximeter, but not in the case of the blood pressure monitor. The majority of participants were unable to read their blood pressure, indicating that they were unable to interpret it. Thus, participants can use the device correctly, but they need help to understand and interpret their physiological data. Understanding and interpreting data would require more knowledge.

In light of the above observations, our results appear to validate our hypothesis (Figure 3). Knowledge has an impact on *effectiveness* (H1) and partially on *satisfaction* (H3). However, the second hypothesis (H2) concerning the link between health knowledge and the handling time (efficiency) of these two devices could not be validated because no significant link could be observed.

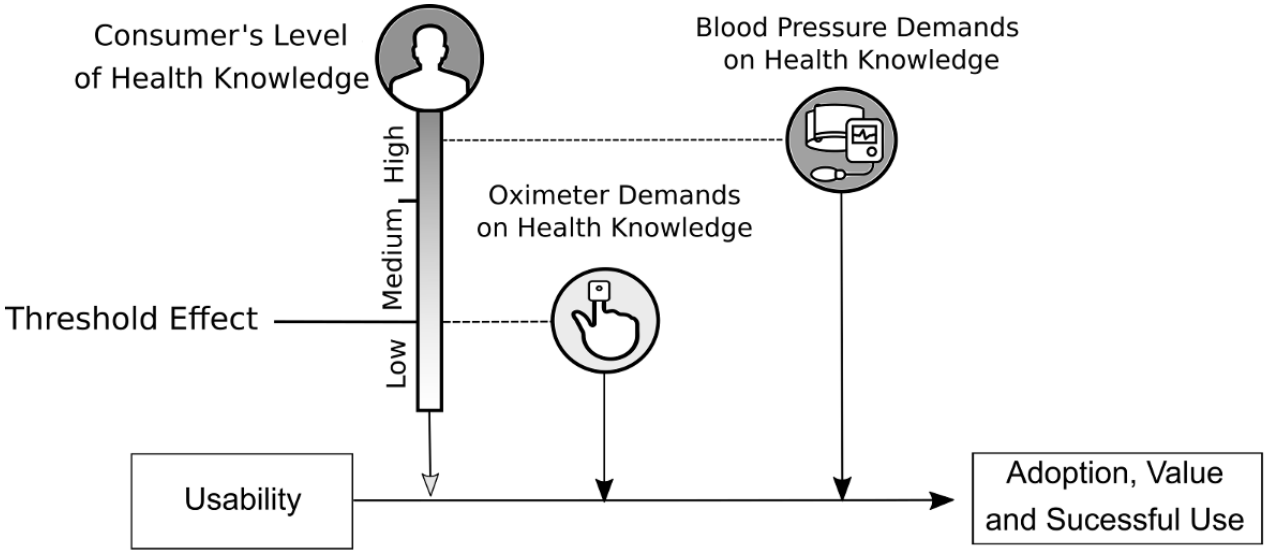
Figure 3. Synthesis of the results of the influence of prior health knowledge on usability results (ISO 9241-11) [15]. Results shown in red correspond to the blood pressure monitor, and those in blue correspond to the pulse oximeter.



* Correlation is significant at the 0.05 level.
** Correlation is significant at the 0.01 level.

These results may be explained by adapting the model from Monkman and Kushniruk [53] by switching the literacy level to the subject's level of knowledge. The oximeter requires a lower level of knowledge on the part of the subject. As for the blood pressure monitor, it is worn on the wrist, which contradicts a widespread belief that blood pressure can only be monitored via the arm. The knowledge related to this device thus demands a greater effort to understand the operation of a blood pressure monitor and involves deeper knowledge. According to the Monkman and Kushniruk model, if a device's "demands on eHealth literacy" exceed "consumers' levels of eHealth literacy," the adoption of the device is compromised. In this framework, the limit of understanding the pulse oximeter is likely to be located between the *low* group and the *medium* group, which explains the threshold effect (Figure 4). This interpretation is also supported by Paasche-Orlow and Wolf [57], who previously observed a threshold effect between health literacy and health outcomes. In contrast, the blood pressure monitor, unlike the pulse oximeter, requires a high level of knowledge in its use and in the reading of its results.

Figure 4. Adaptation of the Monkman and Kushniruk model [52] to the use of the blood pressure monitor and pulse oximeter.



According to the Monkman and Kushniruk model [52], *adoption*, *value*, and *successful use* are related to both the usability and utility fields. However, one of the limitations of our study is that we performed this experiment with a young healthy population that had little interest in using HMDs. As a result, motivation may have affected our outcomes. For O'Brien and Toms [34], usability was linked to the engagement experience. A lack of participant motivation could give lower measures of system usability. In addition, only the performance of participants who successfully collected their health data with the devices was analyzed. If the participants who failed to collect their data had been included, the error count measures would certainly have increased and the SUS score would have decreased.

Another limitation of our study is that the health knowledge questionnaire was designed for the sort of distribution of knowledge that is to be found among a population of university students. However, in the overall population, the distribution is likely much wider than in a university population. This sample is not representative of the national population. For instance, this sample does not include individuals with language, reading, and writing difficulties. This poorly understood link between prior health knowledge and usability could be further explored in the future and especially among populations that are more representative of end users, such as populations that are older and have pathologies. Finally, the choice of equipment was made by health professionals who were accustomed to handling medical equipment. They judged for themselves the simplicity of the equipment, even though it was apparently complex to use. This observation highlights the importance of testing this type of equipment among novice users, which promotes universal design [74], and raises questions about the legislation on European conformity. It is possible that some devices are simpler to use than those chosen in this study. We encourage studies on medical devices that are already on the market in order to investigate possible difficulties in use by the general public [75].

Research Perspectives

This is the first study aimed at detecting the first usability problems outside the context of use, which adds even more complexity. Indeed, the patient will be in a postsurgical context, with pain, nausea, and stress, and will often need the help of family members to use the devices at home. This first study of the *Smart Angel* project serves as a basis for comparing usability measurements on other specific contexts manipulating the four components of usability (*users*, *task* and *goal*, *environment*, and *resources*). A first research perspective is employed to define how the context of ambulatory surgery impacts the usability results of the system. If a patient with a high level of health knowledge makes use errors with a device, the learning context including pain and/or stress may be the cause of this error and therefore a poor usability result.

A second research perspective focuses on the acceptability of the device among patients and medical staff. During the acceptability phase, researchers will focus on the impact of the implementation of the system on the patient pathway and the organization of the hospital. To this end, a hospital study is being carried out in an outpatient surgery population. There is a need to further increase our knowledge of ergonomics on the factors that influence acceptability in order to improve patient safety.

Recommendations for the Design and Integration of HMDs

Three types of recommendations based on the results obtained can be suggested. First, the results of this study could be used in the hospital to improve patient monitoring involving medical devices and to avoid use errors. They could help determine which devices would be the most suitable for the individual's profile in terms of health understanding [76]. This information on a patient's health knowledge would assist physicians in deciding if they need to recommend a particular device to that patient. However, this would require further studies on a population more representative of the national population.

Second, we were able to observe from our results that specific knowledge of the human body (eg, heart function, blood

pressure, and blood oxygenation) made it possible to reduce the number of use errors, even though the individual had no experience with medical devices. Therefore, therapeutic education based on human body functioning, diseases, how to care for devices, and how devices work to meet the patient's needs would be the key to making HMDs more usable [55]. Health professionals could provide anatomical and physiological explanations adapted to the HMDs chosen in terms of body functioning among patients, which will help in taking measurements correctly and preserving patient safety.

Third, we recommend that designers pay attention to the terms chosen. It is important to have a rigorous methodology on usability design and to follow the guidelines relating to the use of clear simple language in health care communication [77]. Just like medical professionals, designers can add playful information about the functioning of the human body to the device instructions.

Conclusion

Our study has two main contributions. First, a scientific interest to provide theoretical knowledge about the factors influencing usability. Indeed, our findings indicate that prior knowledge influences the effectiveness of HMDs.

Second, our study has an applied interest to help designers and medical staff target the importance of providing specific knowledge of the subject to help patients understand how the device works. The results show that it is possible for HMDs to be well adapted to a low literacy level among patients. This was the case, for example, with the oximeter used in this study. Participants were not familiar with this device, and yet, they were able to use it and read their health results.

It should be noted that the study had some limitations. Our sample was restricted to younger adults with high levels of education and adequate health literacy. Clearly, usability needs to be evaluated in larger and more diverse user groups.

Acknowledgments

This project has been financially supported by the European Regional Development Fund, Evolucare, and the *Programme d'Investissements d'Avenir* (PIA), which is part of the PSPC 5 *Projets de Recherche et Développement Structurants pour la Compétitivité*. We would like to thank Elisabeth Chaniaud for her help with the health questionnaire.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The French health knowledge questionnaire.
[DOCX File , 20 KB - mhealth_v8i9e17983_app1.docx]

Multimedia Appendix 2
English translation of the prior health knowledge questionnaire.
[DOCX File , 21 KB - mhealth_v8i9e17983_app2.docx]

Multimedia Appendix 3
Details of user characteristics based on usability measurements for the blood pressure monitor and the pulse oximeter.
[DOCX File , 23 KB - mhealth_v8i9e17983_app3.docx]

References

1. National Research Council. Safe, Comfortable, Attractive, and Easy to Use: Improving the Usability of Home Medical Devices. Washington, DC: National Academies Press; 1996.
2. Bitterman N. Design of medical devices--a home perspective. Eur J Intern Med 2011 Feb;22(1):39-42. [doi: 10.1016/j.ejim.2010.09.017] [Medline: 21238891]
3. Huvila I, Enwald H, Eriksson-Backa K, Hirvonen N, Känsäkoski H. Taking Health Information Behaviour into Account in the design of e-health services. Finnish Journal of EHealth and EWelfare 2016;8(4):153-163 [FREE Full text]
4. Yu DX, Parmanto B, Dicianno BE, Watzlaf VJ, Seelman KD. Accessibility needs and challenges of a mHealth system for patients with dexterity impairments. Disabil Rehabil Assist Technol 2017 Jan 7;12(1):56-64. [doi: 10.3109/17483107.2015.1063171] [Medline: 26153097]
5. Mykityshyn AL, Fisk AD, Rogers WA. Learning to use a home medical device: mediating age-related differences with training. Hum Factors 2002;44(3):354-364. [doi: 10.1518/0018720024497727] [Medline: 12502154]
6. Kortum P, Peres SC. Evaluation of Home Health Care Devices: Remote Usability Assessment. JMIR Hum Factors 2015 Jun 05;2(1):e10 [FREE Full text] [doi: 10.2196/humanfactors.4570] [Medline: 27025664]
7. Carayon P, Hoonakker P. Human Factors and Usability for Health Information Technology: Old and New Challenges. Yearb Med Inform 2019 Aug;28(1):71-77 [FREE Full text] [doi: 10.1055/s-0039-1677907] [Medline: 31419818]

8. Cifter AS. Blood Pressure Monitor Usability Problems Detected Through Human Factors Evaluation. Ergonomics in Design 2017 May 22;25(3):11-19. [doi: 10.1177/1064804617691397]
9. Greysen SR, Chin Garcia C, Sudore RL, Cenzer IS, Covinsky KE. Functional impairment and Internet use among older adults: implications for meaningful use of patient portals. JAMA Intern Med 2014 Jul;174(7):1188-1190 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1864] [Medline: 24839165]
10. Tarricone R, Torbica A, Drummond M. Challenges in the Assessment of Medical Devices: The MedtecHTA Project. Health Econ 2017 Feb 31;26 Suppl 1:5-12. [doi: 10.1002/hec.3469] [Medline: 28139084]
11. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol 2013 Jul 01;7(4):1039-1056 [FREE Full text] [doi: 10.1177/193229681300700429] [Medline: 23911188]
12. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, Implementing, and Evaluating Mobile Health Technologies for Managing Chronic Conditions in Older Adults: A Scoping Review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
13. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
14. IEC 62366-1:2015 Medical devices — Part 1: Application of usability engineering to medical devices. ISO. URL: http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/31/63179.html [accessed 2020-09-08]
15. ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. ISO. URL: http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/35/63500.html [accessed 2020-09-08]
16. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
17. Lyles CR, Sarkar U, Osborn CY. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. Curr Diab Rep 2014 Oct;14(10):534 [FREE Full text] [doi: 10.1007/s11892-014-0534-9] [Medline: 25173689]
18. Peute LW, Driest KF, Marcilly R, Bras Da Costa S, Beuscart-Zephir M, Jaspers MW. A framework for reporting on human factor/usability studies of health information technologies. Stud Health Technol Inform 2013;194:54-60. [Medline: 23941930]
19. Peute LW, Spithoven R, Bakker PJ, Jaspers MW. Usability studies on interactive health information systems; where do we stand? Stud Health Technol Inform 2008;136:327-332. [Medline: 18487752]
20. Tractinsky N. The Usability Construct: A Dead End? Human–Computer Interaction 2017 May 02;33(2):131-177. [doi: 10.1080/07370024.2017.1298038]
21. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012 May;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
22. Borsci S, Federici S, Malizia A, De Filippis ML. Shaking the usability tree: why usability is not a dead end, and a constructive way forward. Behaviour & Information Technology 2018 Nov 02;38(5):519-532. [doi: 10.1080/0144929x.2018.1541255]
23. Bevan N, Carter J, Harker S. ISO 9241-11 Revised: What Have We Learnt About Usability Since 1998? In: Human-Computer Interaction: Design and Evaluation. HCI 2015. Lecture Notes in Computer Science. 2015 Presented at: International Conference on Human-Computer Interaction; August 2-7, 2015; Los Angeles, CA, USA p. 143-151. [doi: 10.1007/978-3-319-20901-2_13]
24. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
25. Sonderegger A, Schmutz S, Sauer J. The influence of age in usability testing. Appl Ergon 2016 Jan;52:291-300. [doi: 10.1016/j.apergo.2015.06.012] [Medline: 26360221]
26. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability Study of Mainstream Wearable Fitness Devices: Feature Analysis and System Usability Scale Evaluation. JMIR Mhealth Uhealth 2018 Nov 08;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]
27. Monkman H, Kushniruk A. A health literacy and usability heuristic evaluation of a mobile consumer health application. Stud Health Technol Inform 2013;192:724-728. [Medline: 23920652]
28. Jensen JD, King AJ, Davis LA, Guntzviller LM. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep;22(6):804-826. [doi: 10.1177/0898264310366161] [Medline: 20495159]
29. van der Vaart R, van Driel D, Pronk K, Paulussen S, Te Boekhorst S, Rosmalen JG, et al. The Role of Age, Education, and Digital Health Literacy in the Usability of Internet-Based Cognitive Behavioral Therapy for Chronic Pain: Mixed Methods Study. JMIR Form Res 2019 Nov 21;3(4):e12883 [FREE Full text] [doi: 10.2196/12883] [Medline: 31750839]
30. Jones J, Caird JK. The Usability of Blood Glucose Meters: Task Performance Differences Between Younger and Older Age Groups. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2017 Sep 28;61(1):604-608. [doi: 10.1177/1541931213601636]

31. Kaufman DR, Patel VL, Hilliman C, Morin PC, Pevzner J, Weinstock RS, et al. Usability in the real world: assessing medical information technologies in patients' homes. J Biomed Inform 2003;36(1-2):45-60 [FREE Full text] [doi: 10.1016/s1532-0464(03)00056-x] [Medline: 14552846]
32. Sparkes J, Valaitis R, McKibbon A. A usability study of patients setting up a cardiac event loop recorder and BlackBerry gateway for remote monitoring at home. Telemed J E Health 2012;18(6):484-490. [doi: 10.1089/tmj.2011.0230] [Medline: 22676379]
33. Harte R, Hall T, Glynn L, Rodríguez-Molinero A, Scharf T, Quinlan LR, et al. Enhancing Home Health Mobile Phone App Usability Through General Smartphone Training: Usability and Learnability Case Study. JMIR Hum Factors 2018 Apr 26;5(2):e18 [FREE Full text] [doi: 10.2196/humanfactors.7718] [Medline: 29699969]
34. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
35. Rico-Olarte C, López D, Kepplinger S. Towards a Conceptual Framework for the Objective Evaluation of User Experience. In: Design, User Experience, and Usability: Theory and Practice. DUXU 2018. Lecture Notes in Computer Science. 2018 Presented at: International Conference on Design, User Experience, and Usability; July 15-20, 2018; Las Vegas p. 546-559. [doi: 10.1007/978-3-319-91797-9_39]
36. Tissier I, Gronier G. Impact of usability, user experience and motivation on the engagement to using a mobile application. In: IHM '14: Proceedings of the 26th Conference on l'Interaction Homme-Machine. 2014 Presented at: 26th Conference on l'Interaction Homme-Machine; October 2014; Luxembourg p. 195-200. [doi: 10.1145/2670444.2670470]
37. 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]
38. Kim H, Xie B. Health literacy in the eHealth era: A systematic review of the literature. Patient Education and Counseling 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
39. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
40. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
41. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low Health Literacy and Health Outcomes: An Updated Systematic Review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
42. Triantafyllidis A, Velardo C, Chantler T, Shah SA, Paton C, Khorshidi R, SUPPORT-HF Investigators. A personalised mobile-based home monitoring system for heart failure: The SUPPORT-HF Study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]
43. Bevan N. International Standards for Usability Should Be More Widely Used. Journal of Usability Studies 2009;4(3):106-113.
44. Frøkjær E, Hertzum M, Hornbæk K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In: CHI '00: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 2000 Presented at: SIGCHI conference on Human Factors in Computing Systems; April, 2000; The Hague, The Netherlands. [doi: 10.1145/332040.332455]
45. Sheehan B, Lucero RJ. Initial Usability and Feasibility Evaluation of a Personal Health Record-Based Self-Management System for Older Adults. EGEMS (Wash DC) 2015;3(2):1152 [FREE Full text] [doi: 10.13063/2327-9214.1152] [Medline: 26290889]
46. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR Mhealth Uhealth 2014 Aug 13;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
47. Al Ayubi SU, Parmanto B, Branch R, Ding D. A Persuasive and Social mHealth Application for Physical Activity: A Usability and Feasibility Study. JMIR Mhealth Uhealth 2014 May 22;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
48. Constantinescu G, Kuffel K, King B, Hodgetts W, Rieger J. Usability testing of an mHealth device for swallowing therapy in head and neck cancer survivors. Health Informatics J 2019 Dec 04;25(4):1373-1382 [FREE Full text] [doi: 10.1177/1460458218766574] [Medline: 29618274]
49. Grebin SZ, Echeveste ME, Magnago PF, Tanure RL, Pulgati FH. [Analytical strategy for assessment of usability of medical devices from the user's perspective: a study of hemodialysis patients]. Cad Saude Publica 2018 Aug 20;34(8):e00074417 [FREE Full text] [doi: 10.1590/0102-311X00074417] [Medline: 30133654]
50. Loorbach N, Karreman J, Steehouder M. Adding Motivational Elements to an Instruction Manual for Seniors: Effects on Usability and Motivation. Technical Communication 2007;54:343-358.
51. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]

52. Monkman H, Kushniruk A. eHealth literacy issues, constructs, models, and methods for health information technology design and evaluation. Knowledge Management & E-Learning: An International Journal 2015;7(4):541-549. [doi: 10.34105/j.kmel.2015.07.036]
53. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
54. Margat A, De Andrade V, Gagnayre R. « Health Literacy » et éducation thérapeutique du patient : Quels rapports conceptuel et méthodologique? Educ Ther Patient/Ther Patient Educ 2014 May 23;6(1):10105. [doi: 10.1051/tpe/2014009]
55. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Arch Intern Med 1998 Jan 26;158(2):166-172. [doi: 10.1001/archinte.158.2.166] [Medline: 9448555]
56. Sarkar U, Fisher L, Schillinger D. Is self-efficacy associated with diabetes self-management across race/ethnicity and health literacy? Diabetes Care 2006 Apr;29(4):823-829. [doi: 10.2337/diacare.29.04.06.dc05-1615] [Medline: 16567822]
57. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]
58. Kickbusch IS. Health literacy: addressing the health and education divide. Health Promot Int 2001 Sep;16(3):289-297. [doi: 10.1093/heapro/16.3.289] [Medline: 11509466]
59. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
60. Huhta A, Hirvonen N, Huotari M. Health Literacy in Web-Based Health Information Environments: Systematic Review of Concepts, Definitions, and Operationalization for Measurement. J Med Internet Res 2018 Dec 19;20(12):e10273 [FREE Full text] [doi: 10.2196/10273] [Medline: 30567690]
61. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
62. Hunt DP. The concept of knowledge and how to measure it. Jnl of Intellectual Capital 2003 Mar;4(1):100-113. [doi: 10.1108/14691930310455414]
63. Agarwal R, Prasad J. A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Information Systems Research 1998 Jun;9(2):204-215. [doi: 10.1287/isre.9.2.204]
64. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan P, Thomas B, Weerdmeester B, McClelland A, editors. Usability Evaluation In Industry. London: Taylor and Francis; 1996:189-194.
65. Lewis J, Sauro J. The Factor Structure of the System Usability Scale. In: Human Centered Design. HCD 2009. Lecture Notes in Computer Science. 2009 Presented at: International Conference on Human Centered Design; 2019; Berlin p. 93-103. [doi: 10.1007/978-3-642-02806-9_12]
66. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
67. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. California, United States: CreateSpace Independent Publishing Platform; 2011.
68. Friard O, Gamba M. BORIS: a free, versatile open‐source event‐logging software for video/audio coding and live observations. Methods Ecol Evol 2016;7(11):1325-1330. [doi: 10.1111/2041-210X.12584]
69. Scott AR, Alore EA, Naik AD, Berger DH, Suliburk JW. Mixed-Methods Analysis of Factors Impacting Use of a Postoperative mHealth App. JMIR Mhealth Uhealth 2017 Feb 08;5(2):e11 [FREE Full text] [doi: 10.2196/mhealth.6728] [Medline: 28179215]
70. Elie C, Colombet I. Méthodes d'estimation de la reproductibilité. Sang Thrombose Vaisseaux 2011;23(3):138-145. [doi: 10.1684/stv.2011.0592]
71. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin 1979;86(2):420-428. [doi: 10.1037/0033-2909.86.2.420]
72. Nunnaly J, Bernstein I. Psychometric Theory. Chicago, IL, USA: McGraw Hill; 1978.
73. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
74. Gibbons MC, Lowry SZ, Patterson ES. Applying Human Factors Principles to Mitigate Usability Issues Related to Embedded Assumptions in Health Information Technology Design. JMIR Hum Factors 2014 Dec 18;1(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.3524] [Medline: 27025349]
75. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
76. Chaniaud N, Loup-Escande E, Megalakaki O. Design, Understanding and Usability Evaluation of Connected Devices in the Field of Health: Contribution of Cognitive Psychology and Ergonomics. In: Proceedings of the 20th Congress of the

International Ergonomics Association (IEA 2018). IEA 2018. Advances in Intelligent Systems and Computing. 2018 Presented at: 20th Congress of the International Ergonomics Association; July 2018; Florence, Italy p. 582-591. [doi: 10.1007/978-3-319-96098-2_72]

77. Ruel J, Alaire C. Communiquer pour tous : Guide pour une information accessible. Santé publique France. 2018. URL: https://www.santepubliquefrance.fr/docs/communiquer-pour-tous-guide-pour-une-information-accessible [accessed 2020-09-08]

Abbreviations

CE: European conformity
HMD: home medical device
ICC: intraclass correlation coefficient
IT: information technology
SUS: system usability scale

Edited by G Eysenbach; submitted 27.01.20; peer-reviewed by J Barre, H Sandhu, T Haggerty, I Gabashvili; comments to author 20.03.20; revised version received 01.04.20; accepted 03.06.20; published 21.09.20.

Please cite as:
Chaniaud N, Métayer N, Megalakaki O, Loup-Escande E
Effect of Prior Health Knowledge on the Usability of Two Home Medical Devices: Usability Study
JMIR Mhealth Uhealth 2020;8(9):e17983
URL: http://mhealth.jmir.org/2020/9/e17983/
doi:10.2196/17983
PMID:32955454

Original Paper

Factors Impacting Clinicians' Adoption of a Clinical Photo Documentation App and its Implications for Clinical Workflows and Quality of Care: Qualitative Case Study

Christine Jacob[1,2], BSc, MSc; Antonio Sanchez-Vazquez[3], PhD; Chris Ivory[3], PhD

[1]Anglia Ruskin University, Cambridge, United Kingdom

[2]University of Applied Sciences Northwestern Switzerland, Brugg, Switzerland

[3]Innovation and Management Practice Research Centre, Anglia Ruskin University, Cambridge, United Kingdom

Corresponding Author:
Christine Jacob, BSc, MSc
Anglia Ruskin University
East Rd
Cambridge, CB1 1PT
United Kingdom
Phone: 41 798769376
Email: christine.k.jacob@gmail.com

Abstract

Background: Mobile health (mHealth) tools have shown promise in clinical photo and wound documentation for their potential to improve workflows, expand access to care, and improve the quality of patient care. However, some barriers to adoption persist.

Objective: This study aims to understand the social, organizational, and technical factors affecting clinicians' adoption of a clinical photo documentation mHealth app and its implications for clinical workflows and quality of care.

Methods: A qualitative case study of a clinical photo and wound documentation app called imitoCam was conducted. The data were collected through 20 in-depth interviews with mHealth providers, clinicians, and medical informatics experts from 8 clinics and hospitals in Switzerland and Germany.

Results: According to the study participants, the use of mHealth in clinical photo and wound documentation provides numerous benefits such as time-saving and efficacy, better patient safety and quality of care, enhanced data security and validation, and better accessibility. The clinical workflow may also improve when the app is a good fit, resulting in better collaboration and transparency, streamlined daily work, clinician empowerment, and improved quality of care. The findings included important factors that may contribute to or hinder adoption. Factors may be related to the material nature of the tool, such as the perceived usefulness, ease of use, interoperability, cost, or security of the app, or social aspects such as personal experience, attitudes, awareness, or culture. Organizational and policy barriers include the available clinical practice infrastructure, workload and resources, the complexity of decision making, training, and ambiguity or lack of regulations. User engagement in the development and implementation process is a vital contributor to the successful adoption of mHealth apps.

Conclusions: The promising potential of mHealth in clinical photo and wound documentation is clear and may enhance clinical workflow and quality of care; however, the factors affecting adoption go beyond the technical features of the tool itself to embrace significant social and organizational elements. Technology providers, clinicians, and decision makers should work together to carefully address any barriers to improve adoption and harness the potential of these tools.

(JMIR Mhealth Uhealth 2020;8(9):e20203) doi:10.2196/20203

KEYWORDS

mHealth; mobile health; telehealth; eHealth; health tech; digital health; user-engagement; dermatology; wound care; mobile phone

Introduction

Background

Mobile health (mHealth) tools are gaining importance in health care as they show promise in several capacities, ranging from efficiencies and time-saving [1,2] to decreasing clinicians' workload and enhancing access to care [3,4]. The use of health apps has also contributed to tackling patients' information needs and making them feel more empowered [5]. The data generated by such tools also help clinicians adapt and customize treatment plans accordingly [6], improving patients' quality of care via personalized treatments [5,6]. Although research shows that clinicians have an overall positive attitude toward mHealth, some barriers to adoption persist [7,8].

In clinical photo documentation and dermatology, research has shown the potential of mHealth tools in managing and preventing skin issues [9-11] and that clinicians and patients recognize their value and are generally willing to use them [12-14].

The Global Observatory of eHealth in the World Health Organization defines mHealth as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [15]. mHealth solutions differ from other information and communication technology apps in the sense that they are typically user-driven, accessible, and affordable [16]; consequently, it is very important to understand better the factors affecting user adoption and the respective implications for workflow and quality of care.

Therefore, this research focuses on understanding the factors affecting clinicians' adoption of mHealth and its implications for clinical practice through a case study of a clinical photo and wound documentation app called imitoCam and its adoption by clinicians in Switzerland and Germany.

Founded in 2016 in Switzerland, imito AG is a clinical photo and wound documentation start-up offering the imitoCam app for medical photo and video documentation and wound measurement. It also supports system interoperability and electronic medical record (EMR) integration. Visuals from the app are presented in Multimedia Appendix 1. The app's key features are explained in Multimedia Appendix 2 and include secure photo documentation; direct patient identification via barcode; measurement of the area (and length, width, and circumference) of wounds and specimens; patient timelines to better understand the case progression; categorization of images to enable photo search; and team collaboration via a chat function, for example, second opinions. The visual in Multimedia Appendix 3 demonstrates seamless integration between the app and existing hospital systems. The wide adoption of the app in 15 hospitals and clinics across Switzerland and Germany at the time of writing this paper made it an ideal candidate to explore mHealth usage and its implications for clinical workflow and quality of patient care.

Objective

This work is a part of a larger study focusing on understanding clinicians' adoption of mHealth; a previous study published earlier presented a more detailed account of our theoretical approach [2]. In this paper, we employ a sociotechnical framework [17] to build a comprehensive analysis of the factors affecting clinicians' adoption of mHealth and its implications for workflow and clinical practice following these 3 main steps:

- Investigating the material aspects of technology and their limitations by identifying the utility and limitations of the app as perceived by the users
- Connecting the material aspects of technology to the tasks it enables and facilitates by highlighting the real constraints to its potential as seen by the user
- Identifying the processes resulting from these affordances and determining the resulting interactions taking place in the organization by identifying the implications for clinical practice and quality of care

The following section explains the research method and how the interview questions and subsequently, the analysis, stemmed from these 3 steps.

Methods

A qualitative paradigm was implemented as it gives priority to "the voices of participants" and the individual and unique "reflexivity of the researcher" [18] and for the rich insights it provides, which can help understand clinicians' individual perceptions in different ways, which cannot be achieved by quantitative methods [19,20].

Data Collection

Data were collected via in-depth, semistructured interviews that were conducted via Skype, Google Hangout, or telecon. Physical artifacts such as screenshots of the app, the devices it can be used on, and examples of user feedback were collected to develop a broader assessment of the studied app [21]. Data collection took place from July 2019 to January 2020, and a total of 20 interviews were conducted with 18 participants working in 8 clinics and hospitals across Switzerland and Germany (2 interviews were preparatory alignment interviews about the tool's features and capabilities). The interviews were conducted via telecon and lasted between 17 min and 90 min, with a median of 35 min. A total of 4 participants sent their responses electronically via email as they did not have the time for a live call. Interviews were conducted and recorded by the first author (CJ) in English. The interview topic guide is available in Multimedia Appendix 4. The research themes and questions were developed in line with the Methodological Guidelines for the Study of Materiality and Affordances by Leonardi to crystallize the focus on the data collected in the interviews [17]. Accordingly, the themes in the interview guide were clustered into 3 categories, starting with an understanding of the tool's utilities and limitations, followed by investigating the technical and social factors affecting adoption and a discussion about organizational and policy factors and implications. The data collection phase continued until an acceptable level of saturation was reached, which was when new data did not generate new insights anymore [20].

Sampling Techniques and Participant Profiles

We used a purposive sample where participants were chosen based on their ability to specify rich and in-depth information about the app and its usage [18,19]. Key informants in imito AG were contacted, and snowball sampling was consequently used to identify suitable participants in partner hospitals and clinics. The key selection criteria were that participants must be clinicians or medical informatics experts in one of the partner hospitals or clinics using the app and must have experienced the app for at least several months. The medical informatics experts had a very good overview of the app and its features, given their access to usage statistics, user feedback, and their constant engagement with clinicians to ensure its successful implementation and sustainability. To minimize the risk of selection bias that might result from the key informants selectively picking users with a positive predisposition toward the app, it was decided that the participants would be asked if they could, in turn, suggest other colleagues who used the app and were willing to participate.

The participants worked in 8 hospitals and clinics across Switzerland and Germany, and the sample consisted of 9 clinicians (one of them was also an imito AG team member), 5 medical informatics experts, and 4 other members from the imito AG team, as shown in Table 1.

Table 1. Sample demographics and characteristics (N=18).

Demographics	Values
Function, n (%)	
Clinicians	9 (50)[a]
Medical informatics experts	5 (28)
Other imito AG team members	4 (22)
Gender, n (%)	
Female	3 (17)
Male	15 (83)
Technological awareness (on a scale of 1-10), mean (SD)	7.5 (2.3)
Health care experience (years), mean (SD)	13.4 (10.4)
mHealth experience (years), mean (SD)	3.9 (2.2)
Location	Switzerland and Germany

[a]One of them is also an imito AG team member.

Data Analysis and Ethical Considerations

Thematic analysis was used to identify and extract the relevant themes and interpret their potential meaning and interrelationships among [19,22]. Computer-assisted qualitative data analysis software, QSR's NVivo was used for data coding. Excerpts were chosen to create an account that expressed the narrative of each theme in a way that helped the reader better understand the analysis. The first author (CJ) conducted the interviews and performed the initial analysis and coding; she is a digital strategist with more than 18 years of experience and has contributed to the creation and realization of several digital tools in health care. The second author (ASV) reviewed the coding; any cases of disagreement were discussed in conjunction with the last author (CI) and mutually agreed upon. The phases of the thematic analysis are clarified in detail in Multimedia Appendix 5 [22].

Our themes were mostly influenced by sociotechnical theory and Leonardi's methodological guidance [17] looking into the technical, social, and organizational factors interacting with shape technology adoption. We also took into account the emerging themes in the most used frameworks for studying mHealth adoption based on a systematic review that we had published earlier [23]. Accordingly, our themes were also influenced by other prominent frameworks such as the Technology Acceptance Model [24], the Diffusion of Innovation theory [25], different forms of extensions of the Unified Theory of Acceptance and Use of Technology [26-29], and the Consolidated Framework for Implementation Research [30,31]. Multimedia Appendix 6 clarifies in detail how each of our themes was influenced by one or more of the existing frameworks and the themes that emerged from the data.

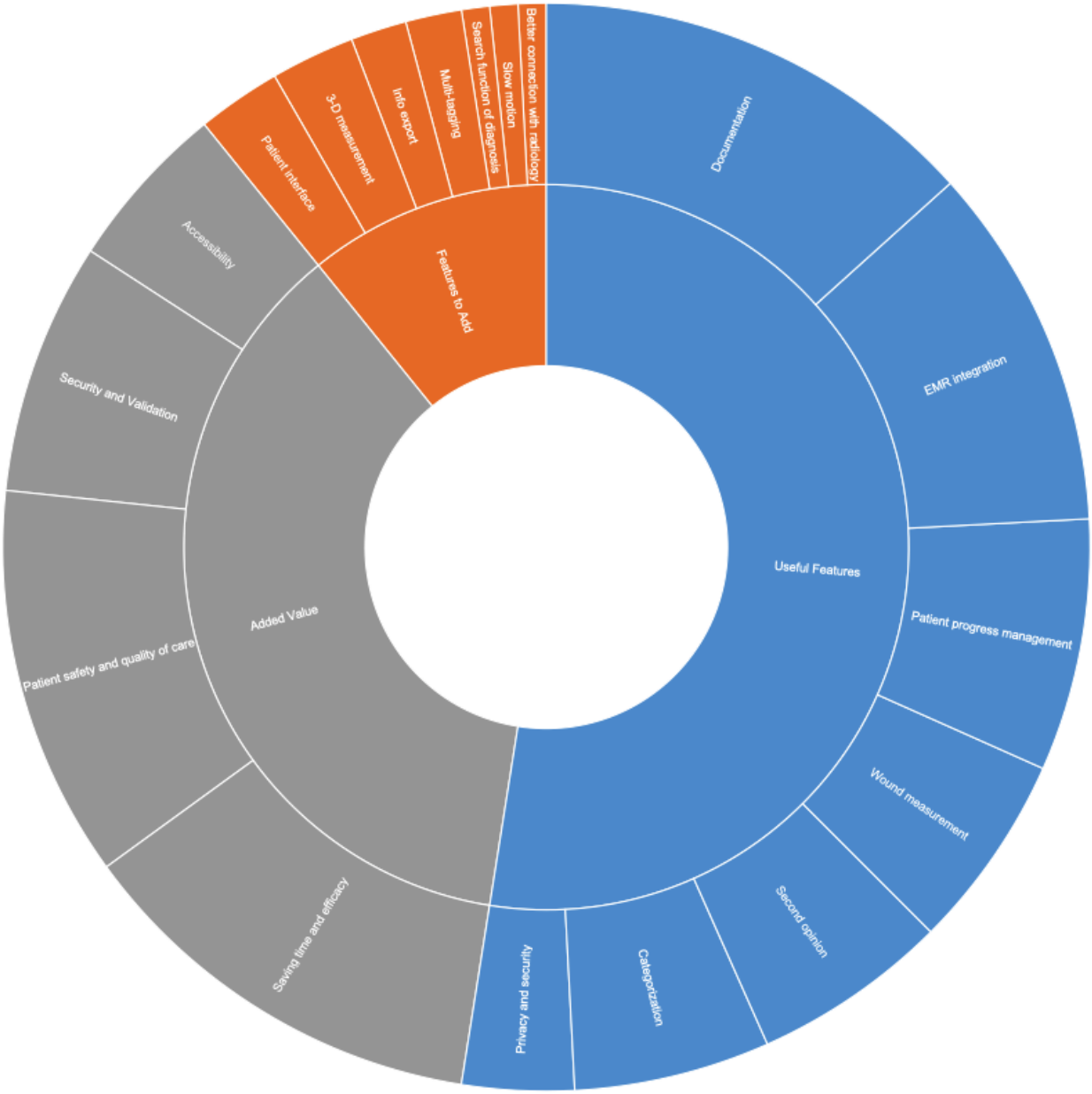
Ethical approval was obtained from the faculty research ethics panel under the terms of Anglia Ruskin University's Research Ethics Policy. All participants were briefed about the research background and signed a consent form agreeing to participate.

Results

Accounting for the Materials: Utility and Limitations

As a first step, we investigated the app's utility and limitations by exploring the most used features, the perceived added value, and the potential limitations or ideas for improvement. Figure 1 shows the themes in these 3 categories and their respective subthemes, reflecting the frequency of each theme (frequencies reflect the number of participants that mentioned that specific theme).

Figure 1. Utility and limitations of the app.



Participants were first asked to name the app features that they used the most to better understand the technological artifacts that they found most useful. Photo and wound documentation (n=16) was the most used feature, followed by EMR integration (n=13), which enables clinicians to link the photos to the right patient in the hospital information system. This was followed by patient progress management (n=9), which gives them visibility of each patient's development over time; wound measurement (n=7) using the imito calibration markers (quick response [QR] codes) for image calibration; and the possibility of obtaining a second opinion (n=7) through the chat function. Participants equally valued the categorization and classification feature (n=7) that enables them to tag the photos and easily find them later. Finally, they also appreciated the privacy and security (n=4) that the app offers as it ensures compliance with the General Data Protection Regulation (GDPR).

Participants were then asked to explain how the app helped them and their patients daily to better understand imito's utility from their perspective. Saving time and efficacy (n=15) was clearly an added value for most of the participants. They also saw an improvement in patient safety and quality of care (n=14) as well as data security and validation (n=9) and explained that the mobility of the app improves data accessibility through the compact overview that it offers (n=6).

To complete the picture, the participants were asked about any limitations they faced or features they would like to add to the app. Limitations mentioned by the participants included the absence of a patient interface, lack of offline functionality, data quality of the user-defined hashtags (eg, typos in user-defined hashtags can limit their searchability), possibility of accessing the imito data from the EMR but not the other way around (eg, the user cannot see the patient data stored in the EMR from the imito tool on their mobile device unless they were captured using the app), the necessity of using physical calibration markers (QR codes) for the measurement function can be cumbersome when the users do not have them on hand, and that

the app can only be used inside the hospital as it requires connection to the hospital's system. Participants also expressed their desire for some additional features such as a patient interface (n=3), 3-dimensional measurement (n=3), export of information from the app via email (n=2), multitagging of body parts or regions to cover injuries involving different body parts (n=2), a search function for diagnoses (n=1), enhancement of the slow-motion video function (n=1), and a better connection with radiology devices (n=1).

The key themes and subthemes, their frequencies, and some sample quotes about utility and limitations are summarized in Table 2 for clarity.

Table 2. Most useful features, added value, and features to add.

Theme	Sample quotes
Most useful features	
Photo and wound documentation (n=16)	• "So that was what we were searching for. A product which is possible to make good photo documentation and the option that it can connect to the system here in the hospital and so we have the picture in the medical file of the patient. And this is the main feature why we use imito because it was the first system that makes it possible in a fast way" [C[a]11]
Electronic medical record integration (n=13)	• "By scanning the name of the patient or his patient identification number, the document is linked to the hospital information system and the photos are stored in the patient file" [C18]
Patient progress management (n=9)	• "very good visibility of the development of each individual case" [P[b]6] • "So, it's easier to follow the progress of healing" [I[c]12] • "Current photos can be immediately compared with older recordings so one can assess the progress of wound healing" [C18]
Wound measurement (n=7)	• "And one further, very good benefit...you can place QR codes in the photo. They are like sticky notes, and you can place it on the screen next to it, and this is referencing it in terms of size. So...you can measure width, length, and even the surface area of a wound. You can decide if a certain area is becoming smaller or larger or whatever" [C5]
Second opinion (n=7)	• "Networking with other authorized users is the next step and enables interdisciplinary communication" [C18]
Categorization and classification (n=7)	• "And before you upload a choice of videos or photos, you are asked to tag your photos by selecting a body region from an illustrated human. And furthermore, you can add hashtags such as 'burn wound'" [C5]
Privacy and security (General Data Protection Regulation compliance; n=4)	• "the only thing today is to send a picture...to other people to get second opinion is using WhatsApp... And it would make my life easier if we would have some good solutions which you are allowed to use, then we could forbid the rest" [I10]
Added value	
Saving time and efficacy (n=15)	• "My expectation was to improve the documentation and make more photos per visit than before, and that certainly worked" [C16]
Patient safety and quality of care (n=14)	• "So, it's an objective parameter, and you see it's getting better. When it's getting better, you continue. If you see it's stagnant, it remains, or it gets bigger, this helps quickly to detect that your medical measures are not good. And then instead of treating the patients another four weeks or three months, you change. You take action and reflect and you change" [C3] • "You can show the patient how his progress is going on and the picture can say more than a lot of words. It's just useful for everyone who's using it" [I12] • "there's also quality benefits that we can directly compare with the initial status" [C16]
Security and validation (n=9)	• "And after the upload, no data is left on the device itself, in the gallery, for example. So, this app sends the images to the hospital's database and there it is as safe as the hospital database can be, and this is the really strong benefit" [C5] • "So, it's not any more than that you have patient pictures that are just flying around somewhere and have no names on it, and you can't map them back to the patient, which was also a matter of patient security and safety" [I13]
Accessibility and compact overview (n=6)	• "It's providing the relevant data at the right time and the right context" [P1] • "Sometimes we need a dermatologist. And so, we can call them. And this is what we want for the whole hospital, that every station is using this for the documentation so that we can sit here in the front of my PC and have a look...So, I don't have to run over there, make the picture, run back or get everything I have, for example, with me. And so, it makes it, for me, easier" [C11]
Features to add	
Patient interface (n=3)	• "So as soon as the electronic health record comes about, then it should be possible to push all that information into the electronic health record of the patients. So, it will be more and more important to let the patient participate on that process" [I13]

Theme	Sample quotes
3-dimensional measurement (n=3)	• "The depth of the wound could also be measured by imito. There was a system that had a laser. And with the laser, you had also the depth of the wound" [C3]
Information export (n=2)	• "And probably, also...the house doctor, and whoever can use the app as an information tool...we have to take the pictures, and then you have to put them in some order, and then you have to export it as a PDF or whatever. That would be great if that could also be mobile and flexible" [C14] • "Once taken, the pictures are imprisoned in imito. You cannot send a GP an email with the photo...So, you are just losing time all the time. You can't reuse the pictures from within imito mobile for presentations because we don't get them out. Whereas respectively, you have to do screenshots. You have to cut the screenshots. You have to send via mail" [C3]
Multitagging body parts (n=2)	• "So right now, you can just choose one body part. And to choose two body parts, this would be an important thing, I think because sometimes we have injuries which are going—they are bigger or just going over different body parts" [I12]
Search function of diagnosis (n=1)	• "I'm missing, namely the long list of diagnoses...and a search function for the diagnosis" [C16]
Better slow-motion video (n=1)	• "it would be beneficial to have a better slow-motion feature in the videos. And I know tools for coaches, for example, golf coaches, and what they can do is while playing in slow motion, they can stop and then they can measure angles, for example" [C5]
Better connection with radiology devices (n=1)	• "we would like to also allow a better connection between those radiological devices...and providing them safe and secure authentication of patients" [CP8]

[a]C: clinician.

[b]P: provider.

[c]I: informatics.

Accounting for Materiality: Constraints and Affordances

We then examined the app's materiality by looking into the constraints and affordances affecting the tool's adoption from technical and social perspectives. Figure 2 shows the themes in these 2 categories and their respective subthemes, reflecting the frequency of each of them.

Technical and material factors evolved around 5 key themes: usefulness, information technology (IT) capability and compatibility, data-related factors, ease of use, and monetary factors. Usefulness is clearly crucial for adoption; most participants explained that the efficacy of the app and the time saved encouraged adoption (n=14), improvements in the quality of patient care also resulted in more usage of the app (n=8), its usefulness in general was valuable (n=5), and the role that it could play in generating scientific evidence via better documentation was also considered (n=2).

The IT capability and integration factors were focused on the app's interoperability (n=13), showing how important it was for the users that the chosen app could integrate effectively with the hospital's local information system, so as to avoid double work and documentation errors, aside from technical issues (n=6) such as poor connectivity, log-in difficulties, or short battery life that are mostly perceived as barriers to adoption. Although data-related factors mostly focused on privacy, security, and data liability issues related to the use of mHealth tools and sharing patient data (n=13), such issues can be overcome with secure and GDPR-compliant tools such as the studied app as it ensures patient data security and privacy; and the challenges related to data management (n=3), especially with the myriad of data generated by such tools which makes the resulting amount of data hard to manage.

Ease of use (n=13) was mostly perceived as a facilitator in the case of imito, with several users mentioning that an easy-to-use tool was central to adoption. Furthermore, monetary factors such as the cost entailed by these tools (n=6) may play a role, not only with regard to the apps' licensing costs but also the related system integration and infrastructure costs. These key technical factors and subthemes, their frequencies, and some sample participant quotes about each of them are summarized in Table 3 for clarity.

Factors affecting adoption went beyond the technical aspects to also cover some social and cultural elements. The users' personal characteristics such as previous experience and habits (n=4), attitude toward technology and change (n=4), and their awareness of the value of such tools (n=3) may also play a role in their decision to adopt such an app. In addition, cultural factors (n=3), such as other people's views and perceptions of using mobile devices at the workplace, may also play a role in the adoption decision. These key social factors and subthemes, their frequencies, and some key participant quotes about each of them are summarized in Table 4.

Figure 2. Technical and social factors impacting user adoption.

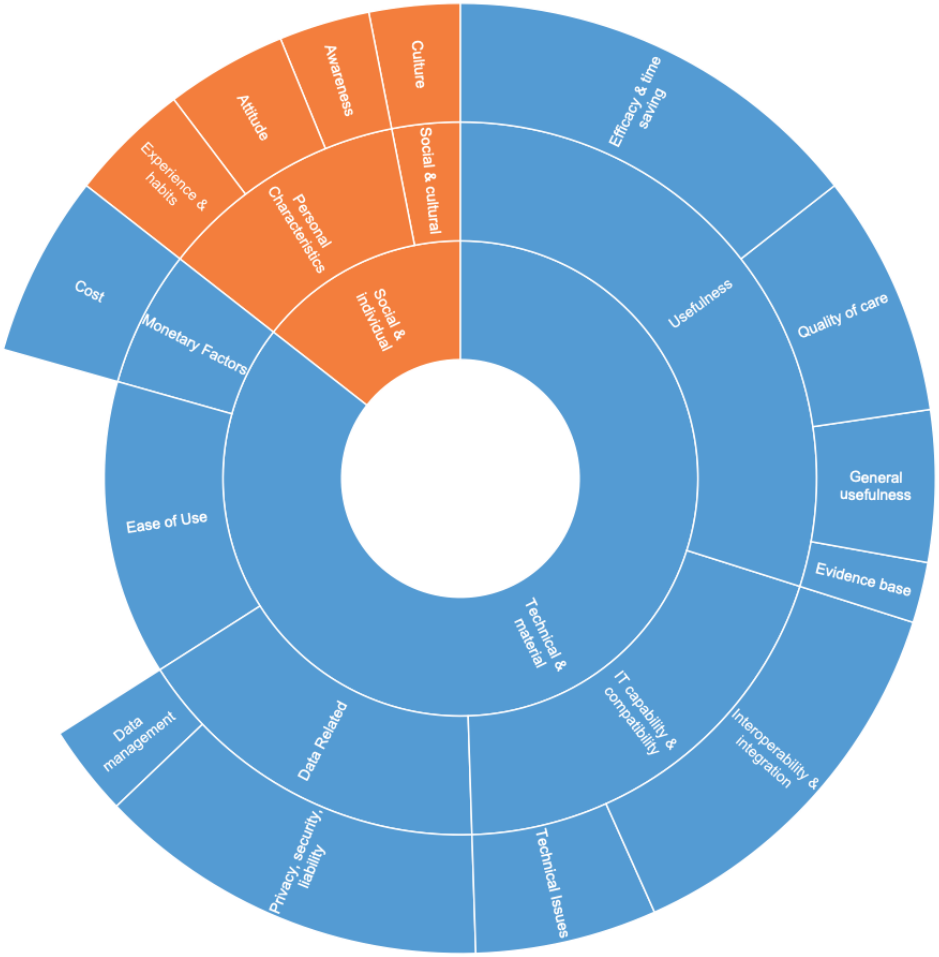


Table 3. Technical factors as expressed by the participants.

Theme	Sample quotes
Usefulness	
Efficacy and time-saving (n=14)	• "It created efficiency. Before it (photo documentation) took maybe three, four, five minutes, and now it takes 30 seconds" [I[a]9] • "So, it is a lot of time-savings and quality improvements" [P[b]15]
Quality of care (n=8)	• "(the case progression overview) helps to quickly detect if your medical measures are not good. And then instead of treating the patients (with the same treatment) another four weeks or three months, you change. You take action and reflect and you change" [C[c]3] • "you take photographs, and you see what are the changes over months or not...(these photos) save time for very specific descriptions that you otherwise place in your report" [C5] • "In the operating room, the photos are not available. So, the clinician has to either just have a good guess what happened in his memory or get to retrieve the photo somewhere else. (With the app) there is really benefits in the treatments because you have the things available when you need them" [P15]
General usefulness (n=5)	• "The clear benefit for the clinical routine" [P1] • "The aspect of creating new possibilities that didn't exist before" [P15]
Evidence base (n=2)	• "I think the more people use the app, we have to see whether it's good for statistics or identifying relevant cases in terms of research" [C5] • "We also expected benefits in terms of scientific studies, simplification of treatment algorithms, and networking of inpatient and outpatient treatment pathways" [C18]
IT[d] capability and compatibility	
Interoperability and integration (n=13)	• "But there are barriers, mainly the IT integration requirement" [P1] • "You can access the app via any mobile device, and logging on with your personal hospital account is possible. The app is then linked to the hospital's database and allows to identify patients by entering their personal details or to scan a barcode and this will give you the patient" [C5] • "The EMR integration in this regard is a challenge both from a cost perspective and the support availability perspective" [I13] • "(the app) is much easier than taking an individual camera as it's directly available within the patient file which is very useful" [C16]
Technical issues (n=6)	• "of course, an app like this needs a lot of battery. So, we have to load the battery two or three times a day" [C11] • "And sometimes, but this is not a problem of the imito app, it's a problem of the system here, when we have no Wi-Fi, it gets more difficult to make a documentation and to save it" [C16] • "I was too frustrated with the log-in process and now that we have the possibility to log in with face ID, it has proved to be a marvel" [C16]
Data-related factors	
Privacy, security, and liability (n=13)	• "The limitations and problems are rather in the legal area, as the sending of sensitive patient data is very restrictive in Germany. Legal and technical requirements for secure data transfer must be dealt with. Good photo documentation supports the sociomedical and legal issues" [C18] • "And altogether you just have to still follow the hospital rules about data security and all that stuff, so that's also an adoption factor. The data security" [I12]
Data management (n=3)	• "And the other thing we noticed is that it needs some kind of controlling in the future because it is so well accepted that some users overdo it. And we are not limited in terms of data capacity, storage space" [C5] • "We have more pictures in this time we roll out the devices. So, I don't know if it's always good to have just more content, if it's also in the right context and is it useful and that stuff. But we have more" [I9]
Ease of use (n=13)	• "I would say the process has to be very easy. So, when you want to have an app like this, it has to be easy, fast, and secure" [I9] • "It's very important to have an easy self-explanatory tool for nurses to use. Otherwise, they won't do it, understandably" [C14]
Monetary factors (cost; n=6)	• "We have a cost in this technological interoperability" [CP8] • "Barriers for establishing such tools are the investment costs, eg, set up of a secure WLAN, equipment, and licensing cost" [C18]

[a]P: provider.

[b]I: informatics.

[a]C: clinician.

[d]IT: information technology.

Table 4. Social factors as expressed by the participants.

Theme	Sample quotes
Personal characteristics	
Experience and habits (n=4)	• "the medical field, as well, has a new generation now, getting to work more with digital health like a tablet or a smartphone" [C[a]11] • "And then the head of the dialysis found out that she really had people on her staff that didn't have a smartphone. But I think it's not the general population in this ward" [C14]
Attitude (n=4)	• "And now with electronic health record opening all of it come these changes that can be challenging for physicians that were not used to that or that are resistant to changes" [CP[b]8]
Awareness (n=3)	• "It's more of an awareness and training topic than functionality...to take a picture, that's very easy, you are used from your own cell phone. But if you make a wound measurement, okay, how does it work? And the QR code and—you have to have some information about this" [I[C]9]
Social and cultural factors (n=3)	• "Maybe on this point of view that, if you ever have a phone in your hands, many people think, 'Okay. You are gaming something, or you are on social media.' But this is a working device. And we are in a change now that the patients—they see, 'Okay. I can do something with the doctor'" [I9]

[a]C: clinician.

[b]P: provider.

[b]I: informatics.

Accounting for Materialization: Organizational Factors and Their Implications

The organizational factors and their implications for clinical workflow and quality of care were then discussed in detail. Figure 3 shows the themes in this category and their respective subthemes, reflecting the frequency of each theme.

Organizational and policy factors revolved around 5 key themes: workflow-related themes, organization's specific inner setting, patient-related factors, user engagement, and policy and regulations. Workflow-related themes were mostly focused on workflow fit and location flexibility (n=12), showing that the app's fit with clinicians' existing work practices encouraged its adoption. Improvements in collaboration and transparency may also increase usage (n=12), whereas traditional clinical practices and infrastructure, such as the lack of use of mobile devices in hospitals, could pose a challenge (n=8). Users naturally favor apps that make their daily work easier (n=6), although existing high workload or lack of resources could sometimes be a barrier (n=5). Some participants explained that the data availability and accessibility facilitated by the app empowered them on the job (n=3). However, apps could also alter some existing roles and responsibilities (n=1). Apps such as imitoCam, for example, have the potential to reduce or even eliminate the role of a professional photographer at the hospital.

Factors related to the organizational inner setting include the complex nature of the decision-making process in hospitals (n=14). Slow decision making can slow down the adoption decision or even prevent it. Participants also pinpointed as to how mHealth tools are replacing traditional tools such as digital cameras in the case of imito (n=10). The importance of training and education to facilitate usage was also highlighted (n=4). The organization's propensity for innovation and appetite for change may also play a role in the adoption decision (n=3), for example, when the organization desires to be perceived as innovative to better compete with other hospitals and clinics. The possibility of trying and piloting the new tool may also facilitate adoption as it minimizes the risk of a full rollout until users have tried the app (n=2).

Patient-related factors were also central and focused mainly on how the app might impact patient engagement and safety (n=11) and enhance access to care (n=2). Many participants also emphasized the significance of user engagement in the development process (n=8), and several imito team members acknowledged the importance of this factor in the success of the tool, explaining that they realized the importance of onboarding their app into the hospital's system and workflows, rather than the other way around.

Policy and regulations in general (n=3), and reimbursement and funding (n=4) in particular, were equally perceived as essential factors for adoption. These key organizational and policy factors, their subthemes, frequencies, and some representative participant quotes are summarized in Table 5.

Figure 3. Organizational and policy factors impacting user adoption.

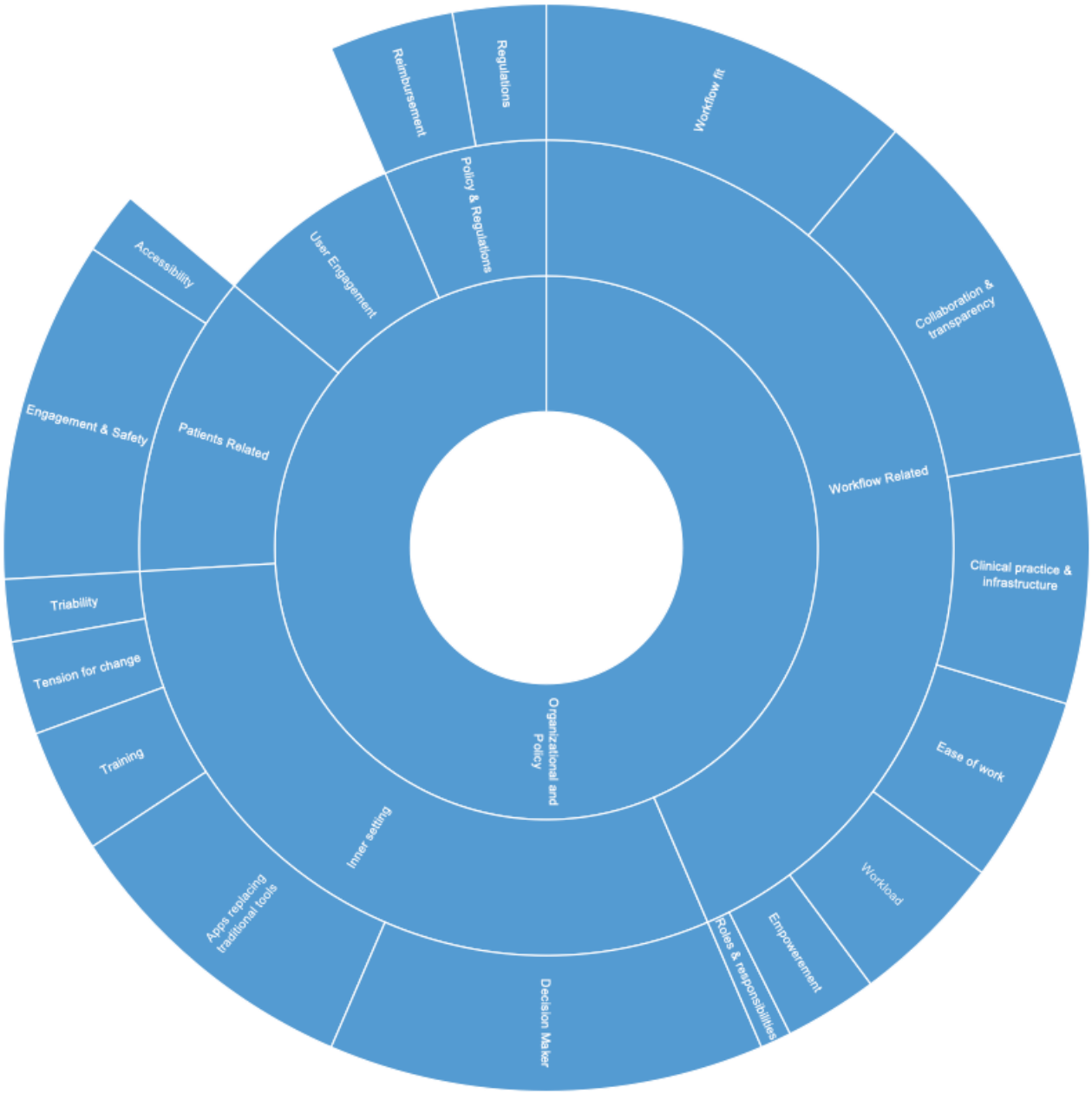


Table 5. Organizational and policy factors as expressed by the participants.

Theme	Sample quotes
Workflow-related theme	
Workflow fit and location flexibility (n=12)	• "before, you had to go onto the station, take the camera. Now, you have it in your pocket right next to you. You can log in with the face ID, take a picture and send it" [I[a]9] • "it's (the app) embedded within the process and the treatment of patients" [I10] • "It's not only the system integration and interoperability but also that workflow integration. So, it helps as a reminder, and it smoothens out the process itself" [C[b]14]
Collaboration and transparency (n=12)	• "For the work on the interdisciplinary team is—it has very good impact. Because, we are working interdisciplinary with surgical dermatologists. And of course, not every time is the surgical physician here; but with the app, we have the possibility here to make a picture and call him" [C11] • "The advantages lie in the improvement of the interdisciplinary cooperation of different medical disciplines and the closer link between inpatient and outpatient treatment pathways" [C18]
Clinical practice and infrastructure (n=8)	• "And in some of the hospitals, it's as well the lacking of mobile devices readiness or how to deal with mobile devices, etc. So, it's more an infrastructure or strategic issue there" [P[c]1]
Ease of work (n=6)	• "It's making the work a lot easier for us" [C11] • "it's easier for the physician to see something in a picture than to read it out of some long description someone did before" [I12]
Workload and resources (n=5)	• "Before taking the decision to adopt we have to check the needed infrastructure for the app. Do we have the technology to roll it out and to use it? And how much work or support does it need to keep on going?" [I12] • "Digitalization is an aid, but it is currently exacerbating the speed and increasing the challenges to performance. It set a much bigger pressure on working forces by creating more demands and increasing speed of everything" [C17]
Empowerment (n=3)	• "You have the power of data so it's a gift in who has the knowledge and often it is used by physician. Physician has the knowledge, has the information in his folder and is coordinating everything, and it gives him big power" [CP8]
Roles and responsibilities (n=1)	• "our professional patient photographer is consulted less frequently, this has changed...it (the app introduction) altered the role of the photographer, it diminished the role a little bit" [C5]
Inner setting	
Decision maker (n=14)	• "decision as it needs quite an intense integration and partnership it will be the IT that makes the decision. But the one that push the decision and that make this decision come through and that is behind the product is really the health care professional" [CP8] • "It took us ages to get through with it. But that was an organizational problem...we had no IT personnel; we had the missing responsibility...we needed buy-in from the local IT guys. And we also need the buy-in from the local MDs of the hospitals, or the managing directors of the hospitals, and so on" [I13] • "It's also one of the barriers, I think. I mean, I'm not totally sure if it was a decision of the ICT department of the medical service...I think it was in connection between the mobility project and the ICT department" [C14] • "I think the problem is nobody's actually willing to make a decision. Everybody wants it. Everybody thinks this is great. But nobody actually says 'Yes. This is going to be implemented'" [P15]
Apps replacing traditional tools (n=10)	• "we obviously wanted to reach more efficiency of daily clinical work because before sometimes you had to find one of the digicams, and they were not that frequent. And it had to be charged, and we needed an SD card. Later on, the SD card had to be brought somewhere else, and he had to store it in an old-fashioned folder system (laughter). So, more efficiency, higher satisfaction for the health care professional itself by more comprehensive documentation" [C5] • "So, the main aspect, the main benefit, is that the manual process that was previously used, I mean, using a point-and-shoot camera and having to transfer the photos from the camera to the computer and then saving them to the right patient. This whole manual process is, yeah, completely replaced by the automatic process. So, it is a lot of time-savings and quality improvements because of the no errors, manual errors, linking the wrong photo to a patient or not linking them at all" [P15]
Training and education (n=4)	• "And when you have high fluctuation of personnel, then you have the problems. You always have to do the training" [C5]

Theme	Sample quotes
Innovation and tension for change (n=3)	• "And I think competition with other health care providers is a topic" [C5] • "the fact we use such an app can also be used in communication, that is something that we use as a tool to also kick off the internal change process in the people and show that (our institution) is an enormous player and open to that kind of innovation" [I13]
Trialability and piloting (n=2)	• "One of the factors is simply pilot projects are available and recommended to take away the fear that something goes wrong" [P7]
Patient-related themes	
Patient engagement and safety (n=11)	• "When you do the things manually at the end of the evening, there's a risk that some picture from patient A go to patient B with the wrong metadata like...I can see it on the picture but is this the right leg or I don't remember and so on. It looks like the left leg. With the app, you do it straight and it's finished and you can work on something else. And the safety and the time is really big thing" [CP8] • "And it's much safer because you have the documentation and you can see it the next time. So, you can compare it with each other" [C11] • "And the cameras, it was always difficulty because you had to go with the SD card to the computer, load it up to the right patient, and the pictures in the SD card, they are not organized. They are just a number, and if you are not watching correctly, you're doing easy mistakes. And in imito, you are more protected from doing these kinds of mistakes" [I12] • "Especially in wound care, they often adapt a treatment because a treatment is not necessarily working. And when they have the photos on the smartphone, they can easily talk to the patient and show them that they can be involved much, much more easily than before because everything is available" [P15]
Accessibility and availability (n=2)	• "The course of healing can be determined by means of photo documentation and information exchange with, eg, outpatient wound care providers and care facilities. For this, the patient does not necessarily have to be presented in the hospital or specialized facility. Unnecessary and long transport routes for patients, eg, from nursing homes are often preventable" [C18]
User engagement (n=8)	• "And then the second is that they realize we're not coming with a solution that we have to onboard the hospital, we do it reverse, we onboard into the hospitals, so they normally stay calm when they realize, aha, you come into our information system, and you work so long until your app works in our system" [P7] • "One of the main parts is the users—so if we have something we think about we could use, we going to show it to the end users and they are pretty much deciding if, in first case, do they actually want it, or do they need it, or they don't" [I12] • "The first thing is that we develop our apps, not on our own. We develop them with the customer. And this really helps to create an app that is made by the customer and for the customer. And then we do a lot of feedback rounds...we go to the customers, to the users, and ask for their feedback and we prioritize" [P15]
Policy and regulations	
Reimbursement and funding (n=4)	• "And there will be no compensation, currently, at least. There will be no compensation for digital solutions, since the federal states are not paying for that...We are working on that. So, we are in close contact with a couple of institutions in the government in order to find some kind of compensation for that kind of expenses" [I13]
Regulations (n=3)	• "...you need a lot of resources and a lot of knowledge to develop a health app. But it's also not so easy to get it through approval, there's a lot of regulations" [C14] • "The limitations and problems are rather in the legal area, as the sending of sensitive patient data is very restrictive...Legal and technical requirements for secure data transfer must be dealt with" [C18]

[a]I: informatics.

[b]P: provider.

[c]C: clinician.

Discussion

Understanding the App's Utilities and Limitations

Participants found the app generally useful, with most users using all the key features and the main utility being efficacy and time-saving as taking clinical photos and documenting them using the app is much quicker and easier. This matches the findings from previous studies that suggested that anticipated improved efficacy enhances the intention to use [2,32-38]. Another utility is the improved patient safety and quality of care by structurally showing each case's progression and enabling the care team to optimize the treatment accordingly; this too validates previous findings that showed mHealth may enable early detection and documentation, resulting in greater safety for patients [39-48].

Clinicians and medical informatics experts were very appreciative of the security and validation aspects of the app, stressing the importance of GDPR compliance and patient data privacy. This utility is of great importance given the typical medicolegal concerns related to confidentiality, inappropriate use, and anonymity of health data [12,45,49-72]. The accessibility and compact overview were clearly an added value of the app as they facilitated timeliness and collaboration, which were also reported in other studies that highlighted how the portability of mHealth tools enabled the care teams to easily access information and flexibly perform tasks anytime and anywhere [1,2,73-76].

Several limitations mentioned by the participants were also reported in previous studies about clinicians' adoption of mHealth, such as potential data quality and management issues [77,78], potential information overload [79-81], and challenges related to data integration and exchange [82-85]. As for the features that the participants wanted added to the app, it was noteworthy to see that some current features such as integration with radiology devices and the offline functionality were on the wish list, revealing that users were not always aware of all available functionalities and underlining the vital role of training and education. Some of the other requested features such as the patient interface are already under consideration by the imito team but are sometimes stalled because of their high development cost.

Understanding Constraints and Affordances

When exploring the factors affecting adoption, users reported several constraints and affordances because of not only technical functionality but also app utility in relation to the particular social and individual context of use. Usefulness was the most prominent technical factor relating to the app's features. In alignment with previous research, perceived usefulness was closely related to time-saving and efficacy resulting from the usage of the app [2,47,48,86], its positive impact on the quality of patient care [66,72,87,88], and the potential benefit for research and scientific evidence because of better data availability [42,89]. Perceived ease of use is an equally important facilitator that has been widely reported in similar studies [2,90-93].

Participants also emphasized the importance of IT factors such as the interoperability and integration of the app with the local system in the hospital or clinic, as this would help them avoid the extra work of having to enter the same data again in the system and what it might entail from documentation errors. This has been perceived as a strong advantage of imito. Interoperability is a known challenge for mHealth and has been reported in many other studies [58,59,94-97]. These are system-level integration issues requiring that the app function properly in relation to existing systems and not just within the bounds of a given mobile device, in addition to other technical issues such as log-in or poor connectivity, which may hinder adoption [64,98-101].

Given the highly regulated nature of health care, factors such as patient data privacy and security are vital for adoption. In the case of imitoCam, this factor was perceived as a facilitator, as the app offers a secure solution for clinical photo documentation, whereas other studies reported this as a barrier if data privacy and security could not be guaranteed [53,59,66,68,82,85,102]. Security requirements thus boost the adoptability of apps that can meet stringent requirements in this regard. Another common challenge relating to data is their management and interpretation, especially when the ease of use of such apps increases the data captured and generated significantly, as discussed in other studies [40,77,78,81,84].

Of course, the cost and cost-benefit play a key role. These may be perceived as facilitators when the app helps in saving costs by creating efficiencies as narrated in similar studies [67,83,103,104]; however, they may also be a barrier [56,105,106] considering the tool's direct costs and the indirect costs related to creating a suitable infrastructure, such as providing handheld devices across the clinic or hospital to support the app's usage.

The findings also revealed that adoption decisions rely not only on the technical and material factors such as app features and the available infrastructure but also embrace some important social and cultural aspects. For instance, the users' individual characteristics such as their previous experience with technology generally, and mHealth specifically, may influence their decision to adopt, as reported by other researchers [1,58,101,107,108]. Their attitudes (eg, resistance to change, risk aversion) may also hinder adoption, and comparable findings were described in earlier studies [39,40,82,109,110]. Even though mHealth is no longer a new concept, cultural views on the use of mobile devices at work may be a barrier [2,111-113]; this is slowly changing and people are accepting these tools more as per our study participants.

Understanding How Technology Materializes in the Organizing Process

The organizational and policy implications of the app's usage were quite prominent, showing that the interaction between the users and the technology creates the adoption patterns that we observe and influences the way people organize their work when using these new tools.

The app's introduction created workflow advantages by offering location flexibility and a better workflow fit as data could be accessed at the point of care and easily embedded in the patient's treatment process—a finding that is aligned with the findings of other studies that a good workflow fit encourages adoption [33,68,91]. Making daily work easier and improving collaboration and transparency were additional workflow advantages described by many participants, as they observed that the app made interdisciplinary teamwork easier and provided more transparency as a direct result of better documentation, which is also in line with the findings of other studies regarding the impact of mHealth on cooperation [12,33,40,67,69,91,103] and streamlining clinical work [114-116]. Empowerment resulting from data availability is another advantage as the app instantly equips clinicians with all the information they need, helping them make more informed decisions, as reported in similar studies [2,100]. Conversely, some other studies stated that mHealth might be perceived as a threat to clinicians' autonomy [43,80,117,118]. We did not find this to be a concern among the participants of this study.

However, workflow disadvantages were apparent around workload and resources. Most clinical staff are already overstretched and the lack of resources may consequently be a barrier to adoption, which has been similarly described in other studies [12,39,92,119,120]. In addition, as one of the participants explained, the greater efficiencies resulting from an app such as imitoCam may also result in a higher workload that could be perceived as an additional burden for clinical staff. The implementation of these new tools may also result in changes in the roles and responsibilities of staff members as reported in other research [45,97,121,122], such as decreasing or eliminating the role of professional photographers as the app replaces their role. Furthermore, the lack of preexisting use of mobile devices in traditional clinical practice is also a challenge as it requires not only workflow adaptations but also infrastructure changes on the part of the hospital.

The nature of the organization of the hospital or clinical setting is also vital for the successful adoption and implementation of mHealth. Factors such as an ambiguous or complicated decision-making process may delay the adoption decision or even prevent it. Other researchers have also shed light on this issue [62,81,92,123], and many participants pinpointed that it is not always easy to identify the people that should be involved in deciding about a new mHealth tool in their organization, and even when they are identified, the process can be quite challenging as the decision involves an interdisciplinary team spanning IT, medical informatics, finance, and medical staff. Given this complexity in decision making, factors such as the trialability of the app and the possibility of piloting it may encourage adoption as it enables decision makers to try a new tool without risking the failure of a broad rollout [48].

Training is another vital aspect of app adoption. This is especially important when the turnover of personnel is high. This point confirms other research that has highlighted the importance of training and education [94,95,124-126]. The organization's desire to be perceived as innovative may also facilitate the adoption of new health apps, a finding that is aligned with those of other studies showing that institutional innovation and openness to change may play a role in the adoption decision [48,127,128].

Patient-related factors reflecting the implications of app usage on quality of care also have an impact on clinicians' adoption of mHealth. Apps that enhance patient engagement and safety, such as imito, have better chances of being adopted. This has also been described in other studies [89,102,129,130]. Similarly, there is a higher acceptance of apps that improve patient access to care [39,67,103]. Furthermore, the external context of the organization, including policy and regulations, may impact adoption. Participants explained that more clarity and simplification of the relevant regulation may facilitate adoption. Specific regulations regarding reimbursement and funding are of special importance for mHealth adoption [52,66,67,84,94,105,131]. The lack or ambiguity of such regulations may hinder the compensation of clinical activities performed via mHealth and may accordingly discourage adoption.

User engagement in the development and implementation of the app is vital for its success [48,54,69,71,99,132-134], and our findings show that this is a mutual responsibility of the health care organization and the mHealth provider. imito offers a best practice example of embedding the partners on the hospital or clinic's side in the development and implementation processes. In this case, they constantly work together with the partner organization (hospital or clinic) in a one-team approach to ensure that the app fits into the technical infrastructure (interoperability and integration) and clinical workflow, taking into account optimization of the clinicians' daily work through the app. This was emphasized in the functionalities implemented to ensure integration with the hospital's systems and processes such as using patient barcodes to identify patients, directly documenting photos in the EMR, and enabling clinicians to log on to the app using the personal code on their hospital badge.

Incorporating user feedback and input in the continuous development and testing of the app is not without challenges, especially when the tool is completely embedded in the hospital's mostly closed system, which results in blocking the visibility of usability statistics from the tool provider. However, imito has established various feedback channels where they get the opinions of the users as well as the inputs of the medical informatics experts that run the tool on the hospital's side to ensure that their app remains relevant and useful.

Limitations and Recommendations for Future Research

This qualitative case study has some limitations that we would like to outline. Our study is limited to a particular mHealth tool and 2 countries in a specific timeframe, and generalization to other settings that might have different characteristics, such as a different regulatory landscape, is not possible. Moreover, the relatively small sample size and the dynamic nature of mHealth necessitate a constant update of the findings to cope with the changes. The sample also excluded nonusers as their recruitment proved to be very challenging, and there is a possible favorable bias in the subgroup of participants who work for the technology provider. Future research may address some of the cited limitations by covering other apps in other countries, timeframes, and settings.

Conclusions

This study demonstrates the utility of the studied app for clinical photo and wound documentation. The app's adoption resulted in several benefits from the participants' perspective, such as time-saving and efficacy, better patient safety and quality of care, data security and validation, and better accessibility. Technical and material factors affecting adoption are usefulness, ease of use, interoperability, cost, and security of the app, whereas social and cultural factors include personal experience, awareness, and attitudes.

Workflow advantages resulting from the app's adoption include better collaboration and transparency, streamlined daily work, clinician empowerment, and improved quality of patient care. Although workflow disadvantages are associated with the available clinical practice infrastructure, workload and resources, the complexity of decision making in hospitals, the need for

continuous training, and the lack or ambiguity of regulations, active user engagement in the development and implementation process may help overcome some of the workflow challenges.

A deeper look into the factors that afford or constrain user acceptance helps us understand materiality at the intersection of social and technical aspects, as the findings show that mHealth adoption is affected not only by the technical features of the app itself but also other social and organizational factors that come into play. These relationships among the technical, social, and organizational factors demonstrate that a successful acceptance and implementation of medical apps not only relies on the tool itself but also necessitates a close collaboration among the tools' providers, clinicians, and decision makers, so that they can address the barriers and harness the potential of these new tools in advancing health care.

Acknowledgments

The authors would like to thank the imito AG team for their openness and collaboration as well as the busy clinicians and medical informatics experts who agreed to give their time and opinions for the sake of this research; this work would not have been possible without their valued inputs.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Visual of imitoCam app.
[PDF File (Adobe PDF File), 5316 KB - mhealth_v8i9e20203_app1.pdf]

Multimedia Appendix 2
Key features of imitoCam at the time of writing this paper.
[PDF File (Adobe PDF File), 1303 KB - mhealth_v8i9e20203_app2.pdf]

Multimedia Appendix 3
Integration of imitoCam into hospital information systems.
[PDF File (Adobe PDF File), 587 KB - mhealth_v8i9e20203_app3.pdf]

Multimedia Appendix 4
Interview guide.
[PDF File (Adobe PDF File), 23 KB - mhealth_v8i9e20203_app4.pdf]

Multimedia Appendix 5
Phases of thematic analysis.
[PDF File (Adobe PDF File), 48 KB - mhealth_v8i9e20203_app5.pdf]

Multimedia Appendix 6
Researcher-driven versus data-driven coding scheme.
[PDF File (Adobe PDF File), 90 KB - mhealth_v8i9e20203_app6.pdf]

References

1. Puszka S, Dingwall KM, Sweet M, Nagel T. E-mental health innovations for Aboriginal and Torres strait islander Australians: a qualitative study of implementation needs in health services. JMIR Ment Health 2016 Sep 19;3(3):e43 [FREE Full text] [doi: 10.2196/mental.5837] [Medline: 27644259]
2. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: qualitative case study. JMIR Mhealth Uhealth 2019 May 3;7(5):e13555 [FREE Full text] [doi: 10.2196/13555] [Medline: 31066710]
3. Kotronoulas G, Papadopoulou C, MacNicol L, Simpson M, Maguire R. Feasibility and acceptability of the use of patient-reported outcome measures (PROMs) in the delivery of nurse-led supportive care to people with colorectal cancer. Eur J Oncol Nurs 2017 Aug;29:115-124. [doi: 10.1016/j.ejon.2017.06.002] [Medline: 28720258]
4. Morilla MD, Sans M, Casasa A, Giménez N. Implementing technology in healthcare: insights from physicians. BMC Med Inform Decis Mak 2017 Jun 27;17(1):92 [FREE Full text] [doi: 10.1186/s12911-017-0489-2] [Medline: 28655299]

5. Kotronoulas G, Connaghan J, Grenfell J, Gupta G, Smith L, Simpson M, et al. Employing patient-reported outcome (PRO) measures to support newly diagnosed patients with melanoma: Feasibility and acceptability of a holistic needs assessment intervention. Eur J Oncol Nurs 2017 Dec;31:59-68. [doi: 10.1016/j.ejon.2017.10.002] [Medline: 29173828]
6. Kotronoulas G, Papadopoulou C, Simpson MF, McPhelim J, Mack L, Maguire R. Using patient-reported outcome measures to deliver enhanced supportive care to people with lung cancer: feasibility and acceptability of a nurse-led consultation model. Support Care Cancer 2018 Nov;26(11):3729-3737. [doi: 10.1007/s00520-018-4234-x] [Medline: 29779057]
7. Connolly SL, Miller CJ, Lindsay JA, Bauer MS. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clin Psychol Sci Pract 2020 Jan 06;27(2):-. [doi: 10.1111/cpsp.12311]
8. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
9. Choi J, Cho Y, Woo H. Mhealth approaches in managing skin cancer: systematic review of evidence-based research using integrative mapping. JMIR Mhealth Uhealth 2018 Aug 2;6(8):e164 [FREE Full text] [doi: 10.2196/mhealth.8554] [Medline: 30072362]
10. Brinker TJ, Schadendorf D, Klode J, Cosgarea I, Rösch A, Jansen P, et al. Photoaging mobile apps as a novel opportunity for melanoma prevention: pilot study. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e101 [FREE Full text] [doi: 10.2196/mhealth.8231] [Medline: 28747297]
11. Brinker TJ, Heckl M, Gatzka M, Heppt MV, Resende Rodrigues H, Schneider S, et al. A skin cancer prevention facial-aging mobile app for secondary schools in Brazil: appearance-focused interventional study. JMIR Mhealth Uhealth 2018 Mar 9;6(3):e60 [FREE Full text] [doi: 10.2196/mhealth.9794] [Medline: 29523502]
12. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to ehealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 5;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
13. Landman A, Emani S, Carlile N, Rosenthal DI, Semakov S, Pallin DJ, et al. A mobile app for securely capturing and transferring clinical images to the electronic health record: description and preliminary usability study. JMIR Mhealth Uhealth 2015 Jan 2;3(1):e1. [doi: 10.2196/mhealth.3481] [Medline: 25565678]
14. Odenheimer S, Goyal D, Jones VG, Rosenblum R, Ho L, Chan AS. Patient acceptance of remote scribing powered by Google glass in outpatient dermatology: cross-sectional study. J Med Internet Res 2018 Jun 21;20(6):e10762 [FREE Full text] [doi: 10.2196/10762] [Medline: 29929947]
15. Kay M, Santos J, Takane M. Mhealth: New Horizons for Health Through Mobile Technologies. World Health Organization. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf? [accessed 2017-06-01]
16. Akter S, Ray P. Mhealth - an ultimate platform to serve the unserved. Yearb Med Inform 2010:94-100. [Medline: 20938579]
17. Leonardi P. Methodological guidelines for the study of materiality and affordances. Routledge Companion to Qual Res Organ Stud 2018:-. [doi: 10.4324/9781315686103-18]
18. Creswell JW, Poth CN. Five qualitative approaches to inquiry. In: Qualitative Inquiry and Research Design: Choosing Among Five Approaches. New York, USA: SAGE Publications; 2018.
19. Clarke V, Braun V. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA: Sage Publications; 2013.
20. Braun V, Clarke V. What can 'thematic analysis' offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
21. Yin R. Case Study Research and Applications: Design and Methods. New York, USA: SAGE Publications; 2017.
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 6;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
24. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
25. Rogers E. Diffusion of Innovations. Fifth Edition. New York, USA: Free Press; 2003.
26. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
27. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Dec Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
28. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
29. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]

30. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
31. Damschroder LJ, Hagedorn HJ. A guiding framework and approach for implementation research in substance use disorders treatment. Psychol Addict Behav 2011 Jun;25(2):194-205. [doi: 10.1037/a0022284] [Medline: 21443291]
32. Charani E, Kyratsis Y, Lawson W, Wickens H, Brannigan ET, Moore LS, et al. An analysis of the development and implementation of a smartphone application for the delivery of antimicrobial prescribing policy: lessons learnt. J Antimicrob Chemother 2013 Apr;68(4):960-967 [FREE Full text] [doi: 10.1093/jac/dks492] [Medline: 23258314]
33. Duhm J, Fleischmann R, Schmidt S, Hupperts H, Brandt SA. Mobile electronic medical records promote workflow: physicians' perspective from a survey. JMIR Mhealth Uhealth 2016 Jun 6;4(2):e70 [FREE Full text] [doi: 10.2196/mhealth.5464] [Medline: 27268720]
34. Putzer GJ, Park Y. Are physicians likely to adopt emerging mobile technologies? Attitudes and innovation factors affecting smartphone use in the Southeastern United States. Perspect Health Inf Manag 2012;9:1b [FREE Full text] [Medline: 22737094]
35. Sandholzer M, Deutsch T, Frese T, Winter A. Predictors of students' self-reported adoption of a smartphone application for medical education in general practice. BMC Med Educ 2015 May 21;15:91 [FREE Full text] [doi: 10.1186/s12909-015-0377-3] [Medline: 25994310]
36. Varsi C, Ekstedt M, Gammon D, Børøsund E, Ruland CM. Middle managers' experiences and role in implementing an interactive tailored patient assessment ehealth intervention in clinical practice. Comput Inform Nurs 2015 Jun;33(6):249-257. [doi: 10.1097/CIN.0000000000000158] [Medline: 25988851]
37. Wilhelmsen M, Høifødt RS, Kolstrup N, Waterloo K, Eisemann M, Chenhall R, et al. Norwegian general practitioners' perspectives on implementation of a guided web-based cognitive behavioral therapy for depression: a qualitative study. J Med Internet Res 2014 Sep 10;16(9):e208. [doi: 10.2196/jmir.3556] [Medline: 25208886]
38. Zhang H, Cocosila M, Archer N. Factors of adoption of mobile information technology by homecare nurses: a technology acceptance model 2 approach. Comput Inform Nurs 2010;28(1):49-56. [doi: 10.1097/NCN.0b013e3181c0474a] [Medline: 19940621]
39. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013 Aug;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]
40. Bramley G, Mangan C, Conroy M. Using telemonitoring to support personal care planning for adults with learning disabilities. J Telemed Telecare 2019 Dec;25(10):602-610. [doi: 10.1177/1357633X18784419] [Medline: 30016895]
41. Giraldo L, Schachner B, Luna D, Benítez S. Exploring nurses' perceptions and expectations toward a BCMA implementation using a mobile app and workstations as a change management strategy. Stud Health Technol Inform 2018;250:134-138. [Medline: 29857405]
42. Kleinpell R, Barden C, Rincon T, McCarthy M, Zapatochny Rufo RJ. Assessing the impact of telemedicine on nursing care in intensive care units. Am J Crit Care 2016 Jan;25(1):e14-e20. [doi: 10.4037/ajcc2016808] [Medline: 26724303]
43. Li L, Cotton A. A systematic review of nurses' perspectives toward the telemedicine intensive care unit: a basis for supporting its future implementation in China? Telemed J E Health 2019 May;25(5):343-350. [doi: 10.1089/tmj.2018.0006] [Medline: 30036155]
44. Moharra M, Almazán C, Decool M, Nilsson A, Allegretti N, Seven M. Implementation of a cross-border health service: physician and pharmacists' opinions from the epSOS project. Fam Pract 2015 Oct;32(5):564-567. [doi: 10.1093/fampra/cmv052] [Medline: 26148726]
45. Penny RA, Bradford NK, Langbecker D. Registered nurse and midwife experiences of using videoconferencing in practice: a systematic review of qualitative studies. J Clin Nurs 2018 Mar;27(5-6):e739-e752. [doi: 10.1111/jocn.14175] [Medline: 29149507]
46. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Perceptions and experiences of heart failure patients and clinicians on the use of mobile phone-based telemonitoring. J Med Internet Res 2012 Feb 10;14(1):e25 [FREE Full text] [doi: 10.2196/jmir.1912] [Medline: 22328237]
47. Steinschaden T, Petersson G, Astrand B. Physicians' attitudes towards eprescribing: a comparative web survey in Austria and Sweden. Inform Prim Care 2009;17(4):241-248. [doi: 10.14236/jhi.v17i4.743] [Medline: 20359402]
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. Hackl WO, Hoerbst A, Duftschmid G, Gall W, Janzek-Hawlat S, Jung M, et al. Crucial factors for the acceptance of a computerized national medication list: insights into findings from the evaluation of the Austrian e-Medikation pilot. Appl Clin Inform 2014;5(2):527-537 [FREE Full text] [doi: 10.4338/ACI-2014-04-RA-0032] [Medline: 25024766]
50. Han KJ, Subramanian R, Cameron GT. Listen before you leap: Sri Lankan health professionals' perspectives on m-health. Health Informatics J 2019 Sep;25(3):858-866 [FREE Full text] [doi: 10.1177/1460458217725903] [Medline: 28825329]

51. Hanna L, May C, Fairhurst K. The place of information and communication technology-mediated consultations in primary care: GPs' perspectives. Fam Pract 2012 Jun;29(3):361-366. [doi: 10.1093/fampra/cmr087] [Medline: 22006040]
52. Hickson R, Talbert J, Thornbury WC, Perin NR, Goodin AJ. Online medical care: the current state of 'eVisits' in acute primary care delivery. Telemed J E Health 2015 Feb;21(2):90-96. [doi: 10.1089/tmj.2014.0022] [Medline: 25474083]
53. Holderried M, Hoeper A, Holderried F, Blumenstock G, Ernst C, Tropitzsch A. Attitudes toward e-health: the otolaryngologists' point of view. Telemed J E Health 2018 Jun;24(6):425-432. [doi: 10.1089/tmj.2017.0158] [Medline: 29227206]
54. Jarvis-Selinger S, Bates J, Araki Y, Lear SA. Internet-based support for cardiovascular disease management. Int J Telemed Appl 2011;2011:342582 [FREE Full text] [doi: 10.1155/2011/342582] [Medline: 21822430]
55. Jetty A, Moore MA, Coffman M, Petterson S, Bazemore A. Rural family physicians are twice as likely to use telehealth as urban family physicians. Telemed J E Health 2018 Apr;24(4):268-276. [doi: 10.1089/tmj.2017.0161] [Medline: 28805545]
56. Jimbo M, Shultz CG, Nease DE, Fetters MD, Power D, Ruffin MT. Perceived barriers and facilitators of using a web-based interactive decision aid for colorectal cancer screening in community practice settings: findings from focus groups with primary care clinicians and medical office staff. J Med Internet Res 2013 Dec 18;15(12):e286 [FREE Full text] [doi: 10.2196/jmir.2914] [Medline: 24351420]
57. Koval PG, Kim JJ, Makhlouf T. Pharmacist perception of a mobile application audience response system for remote pharmacy continuing education participants. J Pharm Pract 2020 Apr;33(2):153-157. [doi: 10.1177/0897190018792391] [Medline: 30099934]
58. Tintorer DL, Domínguez JM, Pujol-Rivera E, Flayeh Beneyto S, Mundet Tuduri X, Saigí-Rubió F. Keys to success of a community of clinical practice in primary care: a qualitative evaluation of the ECOPIH project. BMC Fam Pract 2018 May 9;19(1):56 [FREE Full text] [doi: 10.1186/s12875-018-0739-0] [Medline: 29743030]
59. El Amrani L, Oude Engberink A, Ninot G, Hayot M, Carbonnel F. Connected health devices for health care in French general medicine practice: cross-sectional study. JMIR Mhealth Uhealth 2017 Dec 21;5(12):e193 [FREE Full text] [doi: 10.2196/mhealth.7427] [Medline: 29269336]
60. Mishori R, Anastario M, Naimer K, Varanasi S, Ferdowsian H, Abel D, et al. Mjustice: preliminary development of a mobile app for medical-forensic documentation of sexual violence in low-resource environments and conflict zones. Glob Health Sci Pract 2017 Mar 24;5(1):138-151 [FREE Full text] [doi: 10.9745/GHSP-D-16-00233] [Medline: 28351881]
61. Moskowitz A, Chan YY, Bruns J, Levine SR. Emergency physician and stroke specialist beliefs and expectations regarding telestroke. Stroke 2010 Apr;41(4):805-809 [FREE Full text] [doi: 10.1161/STROKEAHA.109.574137] [Medline: 20167910]
62. Muigg D, Kastner P, Modre-Osprian R, Haluza D, Duftschmid G. Is Austria ready for telemonitoring? A readiness assessment among doctors and patients in the field of diabetes. Stud Health Technol Inform 2018;248:322-329. [Medline: 29726454]
63. Quanbeck A, Gustafson DH, Marsch LA, Chih M, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
64. Radhakrishnan K, Xie B, Berkley A, Kim M. Barriers and facilitators for sustainability of tele-homecare programs: a systematic review. Health Serv Res 2016 Feb;51(1):48-75 [FREE Full text] [doi: 10.1111/1475-6773.12327] [Medline: 26119048]
65. Rogove HJ, McArthur D, Demaerschalk BM, Vespa PM. Barriers to telemedicine: survey of current users in acute care units. Telemed J E Health 2012;18(1):48-53. [doi: 10.1089/tmj.2011.0071] [Medline: 22082107]
66. Anderson K, Francis T, Ibanez-Carrasco F, Globerman J. Physician's perceptions of telemedicine in HIV care provision: a cross-sectional web-based survey. JMIR Public Health Surveill 2017 May 30;3(2):e31 [FREE Full text] [doi: 10.2196/publichealth.6896] [Medline: 28559226]
67. Avey JP, Hobbs RL. Dial in: fostering the use of telebehavioral health services in frontier Alaska. Psychol Serv 2013 Aug;10(3):289-297. [doi: 10.1037/a0028231] [Medline: 22924798]
68. Bailey JV, Tomlinson N, Hobbs LJ, Webster R. Challenges and opportunities in evaluating a digital sexual health intervention in a clinic setting: Staff and patient views. Digit Health 2017;3:2055207617704272 [FREE Full text] [doi: 10.1177/2055207617704272] [Medline: 29942593]
69. 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]
70. Chang F, Paramsothy T, Roche M, Gupta NS. Patient, staff, and clinician perspectives on implementing electronic communications in an interdisciplinary rural family health practice. Prim Health Care Res Dev 2017 Mar;18(2):149-160. [doi: 10.1017/S1463423616000414] [Medline: 27995826]
71. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
72. de Souza CH, Morbeck RA, Steinman M, Hors CP, Bracco MM, Kozasa EH, et al. Barriers and benefits in telemedicine arising between a high-technology hospital service provider and remote public healthcare units: a qualitative study in Brazil. Telemed J E Health 2017 Jun;23(6):527-532. [doi: 10.1089/tmj.2016.0158] [Medline: 27911667]

73. Nerminathan A, Harrison A, Phelps M, Alexander S, Scott KM. Doctors' use of mobile devices in the clinical setting: a mixed methods study. Intern Med J 2017 Mar;47(3):291-298. [doi: 10.1111/imj.13349] [Medline: 27925381]
74. Odnoletkova I, Buysse H, Nobels F, Goderis G, Aertgeerts B, Annemans L, et al. Patient and provider acceptance of telecoaching in type 2 diabetes: a mixed-method study embedded in a randomised clinical trial. BMC Med Inform Decis Mak 2016 Nov 9;16(1):142 [FREE Full text] [doi: 10.1186/s12911-016-0383-3] [Medline: 27825340]
75. Sadoughi F, Erfannia L, Sancholi M, Salmani F, Sarsarshahi A. Teleradiology in southeast Iran: evaluating the views of senior executives and radiologists. Health Care Manag (Frederick) 2017;36(3):301-307. [doi: 10.1097/HCM.0000000000000162] [Medline: 28657915]
76. Sezgin E, Özkan-Yildirim S, Yildirim S. Understanding the perception towards using mHealth applications in practice. Inf Dev 2016 Dec 19;34(2):182-200. [doi: 10.1177/0266666916684180]
77. Brown W, Giguere R, Sheinfil A, Ibitoye M, Balan I, Ho T, et al. Challenges and solutions implementing an SMS text message-based survey CASI and adherence reminders in an international biomedical HIV PrEP study (MTN 017). J Biomed Inform 2018 Apr;80:78-86 [FREE Full text] [doi: 10.1016/j.jbi.2018.02.018] [Medline: 29501908]
78. James S, Perry L, Gallagher R, Lowe J. Diabetes educators: perceived experiences, supports and barriers to use of common diabetes-related technologies. J Diabetes Sci Technol 2016 Sep;10(5):1115-1121 [FREE Full text] [doi: 10.1177/1932296816660326] [Medline: 27417141]
79. Levine M, Richardson JE, Granieri E, Reid MC. Novel telemedicine technologies in geriatric chronic non-cancer pain: primary care providers' perspectives. Pain Med 2014 Feb;15(2):206-213 [FREE Full text] [doi: 10.1111/pme.12323] [Medline: 24341423]
80. MacNeill V, Sanders C, Fitzpatrick R, Hendy J, Barlow J, Knapp M, et al. Experiences of front-line health professionals in the delivery of telehealth: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e401-e407 [FREE Full text] [doi: 10.3399/bjgp14X680485] [Medline: 24982492]
81. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten A. Swedish primary healthcare nurses' perceptions of using digital eHealth services in support of patient self-management. Scand J Caring Sci 2018 Jun;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
82. Bidmead E, Marshall A. A case study of stakeholder perceptions of patient held records: the patients know best (PKB) solution. Digit Health 2016;2:2055207616668431 [FREE Full text] [doi: 10.1177/2055207616668431] [Medline: 29942567]
83. Catan G, Espanha R, Mendes RV, Toren O, Chinitz D. Health information technology implementation - impacts and policy considerations: a comparison between Israel and Portugal. Isr J Health Policy Res 2015;4:41 [FREE Full text] [doi: 10.1186/s13584-015-0040-9] [Medline: 26269740]
84. Chung C, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015 Aug 21;17(8):e203 [FREE Full text] [doi: 10.2196/jmir.4364] [Medline: 26297627]
85. Loh P, Flicker L, Horner B. Attitudes toward information and communication technology (ICT) in residential aged care in Western Australia. J Am Med Dir Assoc 2009 Jul;10(6):408-413. [doi: 10.1016/j.jamda.2009.02.012] [Medline: 19560718]
86. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jan;32(1-2):29-36. [doi: 10.1017/S0266462316000015] [Medline: 27472158]
87. Mueller KJ, Potter AJ, MacKinney AC, Ward MM. Lessons from tele-emergency: improving care quality and health outcomes by expanding support for rural care systems. Health Aff (Millwood) 2014 Feb;33(2):228-234. [doi: 10.1377/hlthaff.2013.1016] [Medline: 24493765]
88. Carlisle K, Warren R. A qualitative case study of telehealth for in-home monitoring to support the management of type 2 diabetes. J Telemed Telecare 2013 Oct;19(7):372-375. [doi: 10.1177/1357633X13506512] [Medline: 24218347]
89. L'Esperance ST, Perry DJ. Assessing advantages and barriers to telemedicine adoption in the practice setting: a MyCareTeam(TM) exemplar. J Am Assoc Nurse Pract 2016 Jun;28(6):311-319. [doi: 10.1002/2327-6924.12280] [Medline: 26178417]
90. Adenuga KI, Iahad NA, Miskon S. Towards reinforcing telemedicine adoption amongst clinicians in Nigeria. Int J Med Inform 2017 Aug;104:84-96. [doi: 10.1016/j.ijmedinf.2017.05.008] [Medline: 28599820]
91. Bello AK, Molzahn AE, Girard LP, Osman MA, Okpechi IG, Glassford J, et al. Patient and provider perspectives on the design and implementation of an electronic consultation system for kidney care delivery in Canada: a focus group study. BMJ Open 2017 Mar 2;7(3):e014784 [FREE Full text] [doi: 10.1136/bmjopen-2016-014784] [Medline: 28255097]
92. Bhatta R, Aryal K, Ellingsen G. Opportunities and challenges of a rural-telemedicine program in Nepal. J Nepal Health Res Counc 2015;13(30):149-153. [Medline: 26744201]
93. de Vries ST, Wong L, Sutcliffe A, Houÿez F, Ruiz CL, Mol PG, IMI Web-RADR Work Package 3b Consortium. Factors influencing the use of a mobile app for reporting adverse drug reactions and receiving safety information: a qualitative study. Drug Saf 2017 May;40(5):443-455 [FREE Full text] [doi: 10.1007/s40264-016-0494-x] [Medline: 28035492]
94. Armstrong AW, Kwong MW, Ledo L, Nesbitt TS, Shewry SL. Practice models and challenges in teledermatology: a study of collective experiences from teledermatologists. PLoS One 2011;6(12):e28687 [FREE Full text] [doi: 10.1371/journal.pone.0028687] [Medline: 22194887]

95. Asua J, Orruño E, Reviriego E, Gagnon MP. Healthcare professional acceptance of telemonitoring for chronic care patients in primary care. BMC Med Inform Decis Mak 2012 Nov 30;12:139 [FREE Full text] [doi: 10.1186/1472-6947-12-139] [Medline: 23194420]
96. Sharma U, Clarke M. Nurses' and community support workers' experience of telehealth: a longitudinal case study. BMC Health Serv Res 2014 Apr 10;14:164 [FREE Full text] [doi: 10.1186/1472-6963-14-164] [Medline: 24721599]
97. Shaw RJ, Kaufman MA, Bosworth HB, Weiner BJ, Zullig LL, Lee SD, et al. Organizational factors associated with readiness to implement and translate a primary care based telemedicine behavioral program to improve blood pressure control: the HTN-IMPROVE study. Implement Sci 2013 Sep 8;8:106 [FREE Full text] [doi: 10.1186/1748-5908-8-106] [Medline: 24010683]
98. Kayyali R, Hesso I, Mahdi A, Hamzat O, Adu A, Gebara SN. Telehealth: misconceptions and experiences of healthcare professionals in England. Int J Pharm Pract 2017 Jun;25(3):203-209. [doi: 10.1111/ijpp.12340] [Medline: 28261891]
99. Molleda L, Bahamon M, St George SM, Perrino T, Estrada Y, Correa Herrera D, et al. Clinic personnel, facilitator, and parent perspectives of ehealth familias unidas in primary care. J Pediatr Health Care 2017;31(3):350-361 [FREE Full text] [doi: 10.1016/j.pedhc.2016.11.001] [Medline: 28012799]
100. O'Connor S, Andrews T. Smartphones and mobile applications (apps) in clinical nursing education: a student perspective. Nurse Educ Today 2018 Oct;69:172-178. [doi: 10.1016/j.nedt.2018.07.013] [Medline: 30096510]
101. Orchard J, Lowres N, Freedman SB, Ladak L, Lee W, Zwar N, et al. Screening for atrial fibrillation during influenza vaccinations by primary care nurses using a smartphone electrocardiograph (iECG): a feasibility study. Eur J Prev Cardiol 2016 Oct;23(2 suppl):13-20. [doi: 10.1177/2047487316670255] [Medline: 27892421]
102. Daniel F, Jabak S, Sasso R, Chamoun Y, Tamim H. Patient-physician communication in the era of mobile phones and social media apps: cross-sectional observational study on Lebanese physicians' perceptions and attitudes. JMIR Med Inform 2018 Apr 6;6(2):e18 [FREE Full text] [doi: 10.2196/medinform.8895] [Medline: 29625955]
103. Armstrong AW, Kwong MW, Chase EP, Ledo L, Nesbitt TS, Shewry SL. Teledermatology operational considerations, challenges, and benefits: the referring providers' perspective. Telemed J E Health 2012 Oct;18(8):580-584. [doi: 10.1089/tmj.2011.0241] [Medline: 22881579]
104. Ayatollahi H, Mirani N, Nazari F, Razavi N. Iranian healthcare professionals' perspectives about factors influencing the use of telemedicine in diabetes management. World J Diabetes 2018 Jun 15;9(6):92-98 [FREE Full text] [doi: 10.4239/wjd.v9.i6.92] [Medline: 29988886]
105. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
106. Jungwirth D, Haluza D. Information and communication technology and the future of healthcare: results of a multi-scenario Delphi survey. Health Informatics J 2019 Mar;25(1):161-173 [FREE Full text] [doi: 10.1177/1460458217704256] [Medline: 28438103]
107. Albrecht U, Afshar K, Illiger K, Becker S, Hartz T, Breil B, et al. Expectancy, usage and acceptance by general practitioners and patients: exploratory results from a study in the German outpatient sector. Digit Health 2017;3:2055207617695135 [FREE Full text] [doi: 10.1177/2055207617695135] [Medline: 29942582]
108. Iacono T, Dissanayake C, Trembath D, Hudry K, Erickson S, Spong J. Family and practitioner perspectives on telehealth for services to young children with autism. Stud Health Technol Inform 2016;231:63-73. [Medline: 27782017]
109. Abd Ghani MK, Jaber MM. Willingness to adopt telemedicine in major Iraqi hospitals: a pilot study. Int J Telemed Appl 2015;2015:136591 [FREE Full text] [doi: 10.1155/2015/136591] [Medline: 26557848]
110. Hines M, Lincoln M, Ramsden R, Martinovich J, Fairweather C. Speech pathologists' perspectives on transitioning to telepractice: what factors promote acceptance? J Telemed Telecare 2015 Dec;21(8):469-473. [doi: 10.1177/1357633X15604555] [Medline: 26377120]
111. Alajlani M, Clarke M. Effect of culture on acceptance of telemedicine in Middle Eastern countries: case study of Jordan and Syria. Telemed J E Health 2013 Apr;19(4):305-311. [doi: 10.1089/tmj.2012.0106] [Medline: 23540280]
112. Farrell M. Use of iPhones by nurses in an acute care setting to improve communication and decision-making processes: qualitative analysis of nurses' perspectives on iPhone use. JMIR Mhealth Uhealth 2016 May 31;4(2):e43 [FREE Full text] [doi: 10.2196/mhealth.5071] [Medline: 27246197]
113. McNally G, Frey R, Crossan M. Nurse manager and student nurse perceptions of the use of personal smartphones or tablets and the adjunct applications, as an educational tool in clinical settings. Nurse Educ Pract 2017 Mar;23:1-7. [doi: 10.1016/j.nepr.2016.12.004] [Medline: 28137514]
114. Schneider T, Panzera AD, Martinasek M, McDermott R, Couluris M, Lindenberger J, et al. Physicians' perceptions of mobile technology for enhancing asthma care for youth. J Child Health Care 2016 Jun;20(2):153-163. [doi: 10.1177/1367493514556555] [Medline: 25427556]
115. Williamson KM, Muckle J. Students' perception of technology use in nursing education. Comput Inform Nurs 2018 Feb;36(2):70-76. [doi: 10.1097/CIN.0000000000000396] [Medline: 29084028]
116. Zilliacus E, Meiser B, Lobb E, Dudding TE, Barlow-Stewart K, Tucker K. The virtual consultation: practitioners' experiences of genetic counseling by videoconferencing in Australia. Telemed J E Health 2010 Apr;16(3):350-357. [doi: 10.1089/tmj.2009.0108] [Medline: 20406122]

117. Liu C, Cheng T. Exploring critical factors influencing physicians' acceptance of mobile electronic medical records based on the dual-factor model: a validation in Taiwan. BMC Med Inform Decis Mak 2015 Feb 7;15:4 [FREE Full text] [doi: 10.1186/s12911-014-0125-3] [Medline: 25889506]
118. Ly BA, Kristjansson E, Labonté R, Bourgeault IL. Determinants of the intention of Senegal's physicians to use telemedicine in their professional activities. Telemed J E Health 2018 Nov;24(11):897-898. [doi: 10.1089/tmj.2017.0276] [Medline: 29470109]
119. Egerton T, Nelligan R, Setchell J, Atkins L, Bennell KL. General practitioners' perspectives on a proposed new model of service delivery for primary care management of knee osteoarthritis: a qualitative study. BMC Fam Pract 2017 Sep 7;18(1):85 [FREE Full text] [doi: 10.1186/s12875-017-0656-7] [Medline: 28882108]
120. Ehrler F, Ducloux P, Wu DT, Lovis C, Blondon K. Acceptance of a mobile application supporting nurses workflow at patient bedside: results from a pilot study. Stud Health Technol Inform 2018;247:506-510. [Medline: 29678012]
121. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]
122. Sturesson L, Groth K. Effects of the digital transformation: qualitative study on the disturbances and limitations of using video visits in outpatient care. J Med Internet Res 2018 Jun 27;20(6):e221 [FREE Full text] [doi: 10.2196/jmir.9866] [Medline: 29950290]
123. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Implementing a telehealth service: nurses' perceptions and experiences. Br J Nurs 2014;23(21):1133-1137. [doi: 10.12968/bjon.2014.23.21.1133] [Medline: 25426527]
124. Dünnebeil S, Sunyaev A, Blohm I, Leimeister JM, Krcmar H. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inform 2012 Nov;81(11):746-760. [doi: 10.1016/j.ijmedinf.2012.02.002] [Medline: 22397989]
125. Duplaga M. Searching for a role of nursing personnel in developing landscape of ehealth: factors determining attitudes toward key patient empowering applications. PLoS One 2016;11(4):e0153173 [FREE Full text] [doi: 10.1371/journal.pone.0153173] [Medline: 27049525]
126. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59 [FREE Full text] [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
127. Casey M, Shaw S, Swinglehurst D. Experiences with online consultation systems in primary care: case study of one early adopter site. Br J Gen Pract 2017 Nov;67(664):e736-e743 [FREE Full text] [doi: 10.3399/bjgp17X693137] [Medline: 28993306]
128. Saigí-Rubió F, Torrent-Sellens J, Jiménez-Zarco A. Drivers of telemedicine use: comparative evidence from samples of Spanish, Colombian and Bolivian physicians. Implement Sci 2014 Oct 8;9:128 [FREE Full text] [doi: 10.1186/s13012-014-0128-6] [Medline: 25293651]
129. Hanley J, Ure J, Pagliari C, Sheikh A, McKinstry B. Experiences of patients and professionals participating in the HITS home blood pressure telemonitoring trial: a qualitative study. BMJ Open 2013 May 28;3(5):- [FREE Full text] [doi: 10.1136/bmjopen-2013-002671] [Medline: 23793649]
130. Miller KE, Kuhn E, Owen JE, Taylor K, Yu JS, Weiss BJ, et al. Clinician perceptions related to the use of the CBT-I coach mobile app. Behav Sleep Med 2019;17(4):481-491. [doi: 10.1080/15402002.2017.1403326] [Medline: 29120247]
131. Choi WS, Park J, Choi JY, Yang J. Stakeholders' resistance to telemedicine with focus on physicians: utilizing the Delphi technique. J Telemed Telecare 2019 Jul;25(6):378-385. [doi: 10.1177/1357633X18775853] [Medline: 29792080]
132. Lord S, Moore SK, Ramsey A, Dinauer S, Johnson K. Implementation of a substance use recovery support mobile phone app in community settings: qualitative study of clinician and staff perspectives of facilitators and barriers. JMIR Ment Health 2016 Jun 28;3(2):e24 [FREE Full text] [doi: 10.2196/mental.4927] [Medline: 27352884]
133. Schmeer R, Behrends M, Kupka T, Meyenburg-Altwarg I, Marschollek M. Use and acceptance of mobile technology by hospital nurses in Germany. Stud Health Technol Inform 2016;225:944-945. [Medline: 27332422]
134. Cox A, Lucas G, Marcu A, Piano M, Grosvenor W, Mold F, et al. Cancer survivors' experience with telehealth: a systematic review and thematic synthesis. J Med Internet Res 2017 Jan 9;19(1):e11 [FREE Full text] [doi: 10.2196/jmir.6575] [Medline: 28069561]

Abbreviations

EMR: electronic medical record
GDPR: General Data Protection Regulation
IT: information technology
mHealth: mobile health
QR: Quick Response

Edited by G Eysenbach; submitted 25.05.20; peer-reviewed by C Varsi, M Lingg, S Six; comments to author 03.07.20; revised version received 05.08.20; accepted 17.08.20; published 23.09.20.

Please cite as:
Jacob C, Sanchez-Vazquez A, Ivory C
Factors Impacting Clinicians' Adoption of a Clinical Photo Documentation App and its Implications for Clinical Workflows and Quality of Care: Qualitative Case Study
JMIR Mhealth Uhealth 2020;8(9):e20203
URL: http://mhealth.jmir.org/2020/9/e20203/
doi:10.2196/20203
PMID:32965232

Original Paper

Smartphone Apps to Support Falls Rehabilitation Exercise: App Development and Usability and Acceptability Study

Helen Hawley-Hague[1,2], MPH, PhD; Carlo Tacconi[3,4], PhD; Sabato Mellone[3,4,5], PhD; Ellen Martinez[6], Mres; Claire Ford[1], BSc; Lorenzo Chiari[3,4,5], PhD; Jorunn Helbostad[7], PhD; Chris Todd[1,2,6,8], PhD

[1]University of Manchester, Manchester, United Kingdom

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

[3]Health Sciences and Technologies-Interdepartmental Center for Industrial Research, University of Bologna, Bologna, Italy

[4]mHealth Technologies s.r.l, Bologna, Italy

[5]Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna, Italy

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

[7]Norwegian University of Science and Technology, Trondheim, Norway

[8]NIHR Applied Research Collaboration Greater Manchester, Manchester, United Kingdom

Corresponding Author:
Helen Hawley-Hague, MPH, PhD
University of Manchester
Division of Nursing, Midwifery and Social Work.
Jean McFarlane Building, Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 7810358568
Email: Helen.Hawley-Hague@manchester.ac.uk

Abstract

Background: Falls have implications for older adults' health and well-being. Strength and balance interventions significantly reduce the risk of falls. However, patients do not always perform the unsupervised home exercise needed for fall reduction.

Objective: This study aims to develop motivational smartphone apps co-designed with health professionals and older adults to support patients to perform exercise proven to aid fall reduction and to explore the apps' usability and acceptability with both health professionals and patients.

Methods: There were 3 phases of app development that included analysis, design, and implementation. For analysis, we examined the literature to establish key app components and had a consultation with 12 older adults attending a strength and balance class, exercise instructors, and 3 fall services. For design, we created prototype apps and conducted 2 patient and public involvement workshops, one with 5 health professionals and the second with 8 older adults from an exercise group. The apps were revised based on the feedback. For implementation, we tested them with one fall service and their patients for 3 weeks. Participatory evaluation was used through testing, semistructured interviews, and focus groups to explore acceptability and usability. Focus groups were conducted with the service that tested the apps and two other services. Qualitative data were analyzed using the framework approach.

Results: On the basis of findings from the literature and consultations in the analysis phase, we selected Behavior Change Techniques, such as goal setting, action planning, and feedback on behavior, to be key parts of the app. We developed goals using familiar icons for patients to select and add while self-reporting exercise and decided to develop 2 apps, one for patients (My Activity Programme) and one for health professionals (Motivate Me). This enabled health professionals to guide patients through the goal-setting process, making it more accessible to nontechnology users. Storyboards were created during the design phase, leading to prototypes of "Motivate Me" and "My Activity Programme." Key changes from the workshops included being able to add more details about the patients' exercise program and a wider selection of goals within "Motivate Me." The overall app design was acceptable to health professionals and older adults. In total, 7 patients and 3 health professionals participated in testing in the implementation phase, with interviews conducted with 6 patients and focus groups, with 3 teams (11 health professionals). Barriers, facilitators, and further functionality were identified for both apps, with 2 cross-cutting themes around phone usability and confidence.

Conclusions: The motivational apps were found to be acceptable for older adults taking part in the design stage and patients and health professionals testing the apps in a clinical setting. User-led design is important to ensure that the apps are usable and acceptable.

(JMIR Mhealth Uhealth 2020;8(9):e15460) doi:10.2196/15460

KEYWORDS

aged; postural balance; telerehabilitation; patient compliance; accidental falls

Introduction

Background

With a rise in the ageing population, there is an increase in the number of falls [1], and around a third of people aged 65 years and above fall each year [1]. Falls have major implications for the health and well-being of older adults. Evidence shows that strength and balance interventions can significantly reduce the risk of falls, rate of falls, and fall injuries [2-5]. However, to be effective, older adults need to reach an adequate level of exercise (thrice a week) and maintain it over time [3]. New innovative digital solutions that support the maintenance of fall prevention exercises and thereby reduce the risk of falls and re-referral to services are needed [6].

The Otago [7] home-based and Falls Management Exercise (FaME) group–based [8] programs are the two main cost-effective strength and balance programs delivered in the United Kingdom [9,10]. However, implementation of these programs by health services often does not conform with evidence-based protocols [9]. Fall rehabilitation is often only delivered over a short period of time, and contact is normally once a week [9]. Patients do not perform the required (unsupervised) home exercises needed to maintain the adequate levels of strength and balance [11] and have poor adherence to the intervention after discharge from rehabilitation services [12,13].

There is emerging evidence supporting the use of mobile phone–based healthy lifestyle programs [14,15], such as those that help increase physical activity [16,17]. King et al [18] developed and tested mobile apps based on the behavior change theory designed to motivate adults aged 45 years and older. One of these included personalized goal setting and behavioral feedback, receiving positive feedback from participants and increasing physical activity. Evidence suggests that mobile phones are more usable than other devices [19]. The proportion of older adults using smartphones is growing rapidly, with 39% of those aged 65 to 74 years and 15% of those aged above 75 years found to be using smartphones [20]. The advantages of smartphones over tablet devices is that the person is more likely to carry it with them and to leave it switched on when not directly using it, which can ensure the delivery of feedback in real time.

If we can use technology to support older adults to maintain an adequate level of strength and balance training, we could support them to maintain health and independence, reduce their fall risk, and prevent re-access to rehabilitation and hospital admissions. Previous research suggests that technologies that support fall prevention are acceptable to older adults as long as they are simple, reliable, effective, and tailored to individual needs [21]. A majority of smartphone apps used for fall prevention focus on risk assessment, rather than support for fall prevention exercise [22]. There is an ongoing trial of a smartphone-based strength and balance app, but this focuses solely on balance exercises [23] and is used on tablet devices. Another app that has been developed in Sweden is designed to be used alone by older adults, rather than in the management of falls as part of rehabilitation with a health professional [24].

Theoretical Approaches

Previous studies have shown that attitudes and beliefs are important to uptake and adherence to exercise by older adults [25,26]. The theory of planned behavior (TPB) [27] is particularly useful for assessing older adults' attitudes in relation to exercise [25,28]. The TPB is based on 3 core components:

1. Perceived behavioral control and perceived ease or difficulty of performing the behavior.
2. Social influences including subjective norms (beliefs of important people, eg, family), perceived social support (support from others for behavior), and modeling (following observed behavior of others).
3. Attitudes [27] focused on the advantages and disadvantages of the behavior (outcome expectations), and when related to adherence, whether these advantages have occurred.

There is evidence that interventions based on TPB can influence a person's intent to exercise [28]. Adherence is also related to attitudes measured by the TPB [25,26].

In psychological literature, goal setting has been found to be a successful behavior change technique [29,30]. Other theories, such as the self-determination theory (SDT), can be related to goal setting and support the process of setting meaningful personal goals [31]. By choosing personal goals, we are more likely to satisfy intrinsic needs of relatedness (a sense of belonging), competence (a feeling of having the skills to achieve goals), and autonomy (that we can take direct action to make a change). From previous research exploring uptake and adherence to exercise programs by older adults, we know that goal setting, outcome-based feedback, and feedback designed to strengthen self-efficacy (eg, praise for progress so far) is particularly important for motivation to exercise [25,28,30,32]. It is important when describing interventions, particularly those based on behavioral theory, that we explore behavior change techniques (BCTs) utilized as mechanisms for this behavior change. The Behavior Change Taxonomy by Michie et al [33] has been developed to allow the BCTs used in an intervention to be clearly described and replicable.

Usability is an important part of designing successful technologies and has been described as the overall usefulness of a product and whether a person can use it for its intended purpose [34]. Acceptability is a multifaceted construct that reflects the extent to which people delivering or receiving a health care intervention consider it appropriate [35]. Smartphone apps are more likely to be acceptable to older adults and health care professionals if they are developed with them [36] using principles of human-centered design (HCD). The HCD ensures that the needs of the user are taken into consideration throughout the design process, and is a multi-stage process that allows for various iterations of a design to ensure it meets the needs of users [36]. Models such as the technology acceptance model (TAM), which focuses on whether a technology is perceived as useful and whether or not it is easy to use [37], are important when developing technologies. This model has also been expanded to include factors such as subjective norms and how this relates to perceived usefulness [38].

Study Aims

The aims of our study are as follows:

1. To develop smartphone apps designed to support patients to exercise, based on psychological theory, and co-designed with health care professionals, older adults, and patients.
2. To explore whether two new smartphone-based apps designed to support adherence are usable and acceptable to health care professionals and patients when supporting a strength and balance home exercise program to prevent falls.

This study includes a range of interacting behavior change components and intervention development and is therefore based on the principles of the Medical Research Council framework for the development and evaluation of complex interventions [39].

Methods

To achieve our aims, a staged approach was undertaken and the system development lifecycle was used as a structure for describing the stages of development and the iterative approach [40]. For this paper, we focus on the first 3 phases: analysis, design, and implementation.

Analysis Phase

During the analysis phase, we wanted to inform the key components of the app through a search of the literature. We had already established that this type of app on smartphones did not exist and instead used the exercise and psychological literature to establish important components of the app.

We then conducted an informal consultation with a community-based strength and balance exercise class of older adults. The 12 older adults (10 women and 2 men) attending the class were asked for their opinions on using an app to set goals and home exercises and then how they would feel about receiving messages and prompts. We also asked how they would feel about reporting their exercises on the phone. Furthermore, the older adults were asked to share examples of the types of outcome-based goals they might set.

Subsequently, we approached exercise instructors delivering the evidence-based program through Later Life Training (LLT). LLT is the largest provider of training to deliver the FaME and Otago program in the United Kingdom [41]. This was done through their Facebook page. We asked instructors what types of goals they set with older adults and the feedback that they gave them during their programs. We also approached 3 fall services (9 physiotherapists, 2 rehabilitation assistants, 2 occupational therapists, and 3 nurses). We asked them what they thought about the initial concept of the app. This included using pictures from exercise booklets provided by LLT, as these are commonly used by falls services in the United Kingdom, examples of goals they could set, and feedback given to patients.

Design Phase

During the second design phase, HH (health care researcher) and CT and SM (computer scientist and engineer, respectively) worked together to put together 2 basic prototype apps based on wireframes. This started by creating diagrams of how the app would flow and the functions, then there were further discussions, and then the creation of the prototypes. As part of this phase, we also looked at the types of BCTs that we had found based on the literature and discussions and mapped them to the Behavior Change Taxonomy by Michie et al [33]. These are reported in detail in our trial protocol paper [42].

Once the initial prototypes were created, 2 patient and public involvement (PPI) app development workshops were conducted to gain initial feedback on the concept (perceived usefulness), basic design (ease of use), and approach developed so far with the following groups:

1. Group of health care professionals (n=5) from a Manchester Falls Service (2 physiotherapists, 1 occupational therapist, 1 rehabilitation assistant, and 1 assistant practitioner).
2. Group of older adults were aged 60 years and above (n=8), community dwelling, and independent living from an age UK strength and balance falls exercise group.

The health care professional workshop was run by a researcher who was also an occupational therapist (OT) in a different fall team. The older adult workshop was run by the OT and lead researcher for the project. In the workshops, the initial concept of the technology was discussed with an explanation of why we thought it was important (perceived usefulness), what we were trying to achieve, and how the apps would work. We connected the phone to a large screen and demonstrated to the group what patients would have to do with the app. For health care professionals, we demonstrated what they and patients would have to do. This was followed by giving participants the opportunity to use the apps themselves. We had several phones, and participants in both workshops took turns to navigate around the apps. The older adults could see examples of exercise programs that had been scheduled and have a go at reporting their exercises. Health care professionals tried goal setting. We discussed participants' thoughts on a one-on-one basis as they tested the apps and then brought everyone back together for further discussion. At this point, feedback messages had not yet been programmed, but potential examples were discussed. Notes were not only taken on the feedback provided but also on observations of participants' use of the phones.

Contact with older adults and health care professionals in these first 2 phases of development was classified as patient and participant involvement (PPI). Therefore, we only collected aggregate details on gender, ethnicity, previous experience of smartphone or tablet use for older adults and gender and clinical background for health care professionals.

Implementation Phase

Usability Study

This stage was used to determine the acceptability and usability of the apps with patients and health care professionals and included participatory evaluation, testing, interviews, and focus groups. The usability study ran for 3 weeks and tested the acceptability and usability of the technology as part of the exercise intervention.

The research proposed in this stage is predominantly qualitative but forms part of a larger mixed methods approach [43]. This approach enables us to establish whether the technology is acceptable to patients and health care professionals (qualitative methods) and assess its usability (technology testing), making improvements if required. The study was granted ethical approval by the North West Greater Manchester Central NHS Ethics Committee. The usability study has allowed for the planning of a subsequent feasibility randomized controlled trial and the design of the apps [42].

Sampling Principles and Procedures

Patients at risk of falls (aged 50 years and older), identified through one community fall rehabilitation service in Manchester, were recruited. As this was a pragmatic study, participants were those who would usually be offered a home exercise program by the service and could be at any stage in their rehabilitation. Patients who were unable to follow instructions were excluded, as were those with severe visual or hearing impairment. At this point, there were no other exclusion criteria. The first 20 eligible patients who were currently attending the service who were willing to participate were recruited. A total of 3 health care professionals gave patients the study information sheet and informed them about the technology. The apps were demonstrated to patients before they were asked to give informed consent. All patients recruited were offered a one-on-one interview in their own home. If they had a family member present, they could also join.

Health care professionals from 3 fall services in Manchester were recruited to participate in 3 focus groups. All members of staff (n=17) in each team were given study information by their team leader and asked if they were available for a focus group, which then took place at their place of work.

The Intervention

For the testing, we used the Samsung Galaxy S4 with pay-as-you-go SIM cards and 4G, and where possible, we connected them to the patients' Wi-Fi. The apps tested on the smartphone included the following: (1) "Motivate Me" app: used by the health care professional in consultation with the patient to set outcome and behavior goals, including scheduling their exercises. The health care professional can then view the patients' exercise reports sent from "My Activity Programme" and send feedback messages. "Motivate Me" sends the patients' exercise program to their "My Activity Programme" app. (2) "My Activity Programme" used by the patient to view their exercise program sent through Motivate Me reports their exercises back to the health care professional and receives prompts to exercise and motivational messages as pop-ups (Multimedia Appendix 1).

The exercise program, goal setting, and feedback were delivered by a health care professional. The health care professional delivered the Otago [7] and FaME [8] exercises. These were delivered once a week face-to-face with a health care professional in the patients' home and they were encouraged to exercise at least three times a week. Exercises delivered on "My Activity Programme" were adapted and tailored to the patient's individual needs based on a pre-exercise assessment already conducted by the service and through goal setting on the app. All patients were given a home exercise booklet, which is part of the standard service.

Testing Procedures and Measurements

Patients were linked to one health care professional to support them with their program and carry out the intervention. This was allocated as would be in standard practice (not influenced by the research team). Issues the health care professional (deliverer) and the patient (recipient) had with the smartphone technology throughout the testing period were recorded (issue log and field notes). We recorded all issues but were particularly focused on perceived usefulness (requirement for internet access or testing of 4G through mobile phone, whether patients received messages) and ease of use (use of touch screen for reporting, ease or acceptability of reporting, whether health care professionals could use the phone to set goals) [37].

Interviews and Focus Groups

Following the testing period, health care professionals from 3 services delivering fall rehabilitation were recruited to participate in 3 focus groups. The service involved in the testing gave direct feedback on their experiences of using the technology. The other 2 services received a demonstration of the technology and were asked to give their feedback based on a similar interview schedule.

Patients who participated took part in a one-on-one interview in their own home. The interview and focus group schedules were based on the FAll Repository for the design of Smart and sElf-adaptive Environments prolonging Independent livinG (FARSEEING) [44] consortium guidelines (Multimedia Appendix 2). Participants demonstrated their use of the phone to the researcher. Key areas were explored in relation to the "Motivate Me" and "My Activity Programme" and can be mapped to either ease of use or perceived usefulness.

Data Analysis

Data from the issue logs were collated and summarized, and comments were added to the qualitative data analysis. Data from the issue logs provided triangulation for focus group or interview data.

Follow-up interviews with patients, focus group data with health care professionals, and field notes were analyzed together using

a framework analysis, where the questions around the different apps provided a natural structure for the coding [45]. The NVivo 11 (QSR International) qualitative data analysis software was used to manage the data. The validity of the analysis was checked by returning to the data once themes were identified and through the use of a second researcher who carried out independent coding. Codes that emerged were discussed between 2 researchers, which is an approach that ensures rigor [46].

Results

Analysis Phase

The initial motivation for the development of the apps came from the literature examining previous studies looking at exercise adherence [13,18,25,26,28,30,32]. We identified the importance of goal setting and setting outcome-based goals (what is it that patients would like to do that they cannot do now) [25,30,32]. We also identified the value of feedback as a mechanism for building self-efficacy and as a means of social support in the exercise and mobile health interventions [14,15,18,32,47].

Health care professionals across the 3 fall services told us that they already carried out goal setting with patients on the first assessment but that this was informal. They indicated that an app that would support this process in a formalized way and providing feedback to patients would enhance current practice.

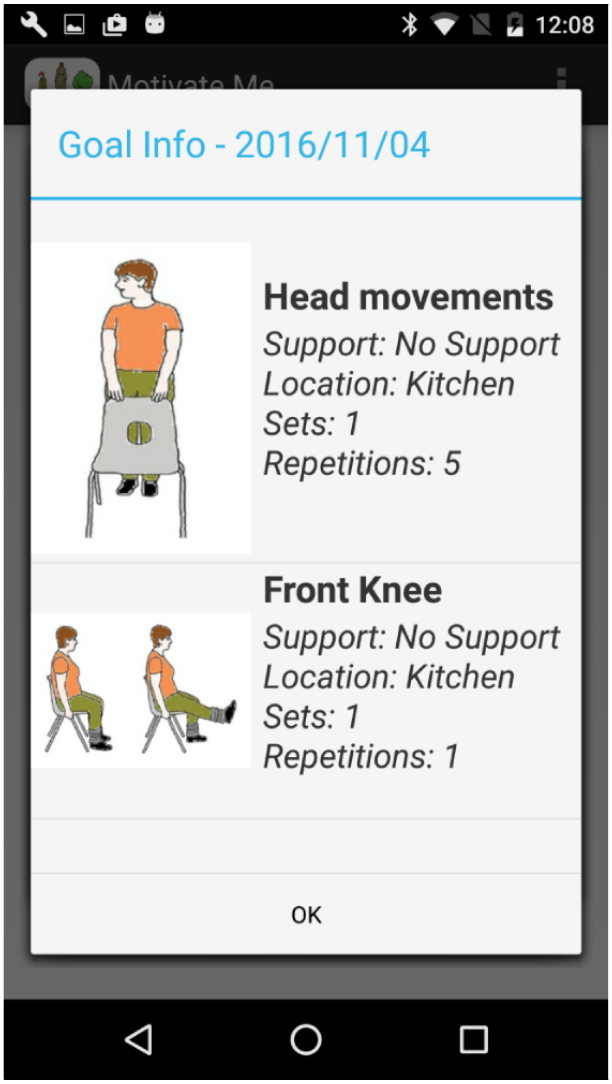
On the basis of the literature [25,29,30,32] and the discussion with health care professionals, we decided to develop long-term outcome-based goals as an important part of the app. Health care professionals usually ask patients about long-term outcome-based goals at the point of first assessment, so we wanted the app to facilitate this. Within the app, we offered a range of preset goals and based on feedback from health care professionals, grouped them into physical (physical ability or task specific) and emotional goals (I want to feel more confident).

The literature suggests that personalization of the exercise program and technology is important to patients [21] and that older adults prefer limited interaction with technology [21]. To ensure that the app was personalized to the individual and responsive to their needs and requirements, we decided to ask health care professionals to do most of the personalization. Therefore, 2 apps were created; one for the health care professional (to carry out goal setting with the patient and to communicate with them) and one for the patient (to receive messages and view exercises or goals). The health care professional could then guide the patient through the goal-setting process and personalize their goals and program. The patients' goals and program would then transfer from their phone to the patients' phone, requiring less interaction and less technical ability from the patient. As sensors in the phone are still unreliable in detecting static exercises without constantly moving the position of the phone or additional sensors [48], we decided not to use automatic detection of exercise through the phone but rather to include self-reporting of exercises within the patient app, which the health care professional could then see. This information provides an indication of the patient's progress and adherence.

We decided to use pictures of the exercises included in the LLT's home exercise booklets (Figure 1). Health care professionals liked the idea of using icons that they and patients were familiar with. We decided to adopt the name "Motivate Me" for the health care professional app. "Motivate Me" is LLT's motivational training for working with older adults and the most adopted motivational training across the United Kingdom [41]. We adopted the name "My Activity Programme" for the patient app because health care professional and older adult feedback suggested we avoided the word "exercise."

BCTs proposed as part of the apps, based on this initial consultation and the evidence [33], include goal setting (behavior and outcome), action planning (recording the plan to exercise in a diary on the smartphone or reminder text messages when it is time to start the program), and feedback on behavior (providing feedback on what they have done or benefits).

Figure 1. Interface: example of exercises.



Design Phase

Basic Prototype

Following this initial consultation and development work that established the potential scope and requirements for the app, we created storyboards and app flow or options for the operationalization of the apps. Prototypes were then created.

In the initial prototypes, both apps had a very simple design, including pale yellow buttons, black writing, and large font to aid visual impairment. No messages were displayed in the initial prototype.

PPI Workshops

Older Adults Workshop

Demographics of participants in both workshops are reported in Multimedia Appendix 3. We observed when we first introduced the concept of the technology that the majority of the older adults (6/8, 75%) were reticent about using smartphones, describing them as frightening. The two who were comfortable with the concept were the participants who had previously used a smartphone or tablet. However, it was observed by the facilitators later on in the group discussion that the older adults having held the phone, looked at the app and found that they were able to use it had some of their fears allayed (6/8, 75% were then positive about using the app). This supports the idea that involving users in the development of new technology means they are more likely to adopt it [21]. Further details of the feedback are outlined in Multimedia Appendix 4. Following these initial workshops, we created "How to" guides for health care professionals and older adults liaising with patients on our study advisory board.

Implementation Phase

In total, 7 patients (4 men; mean age 77.1, SD 8.53 years; range 64-92) took part, 6 agreed to be interviewed, and for one interview, the patient's son was also present. Only two of the patients who took part already owned a smartphone. Further demographics are reported in Multimedia Appendix 1. During the testing, 2 physiotherapists (2 patients each) and 1 occupational therapist (3 patients) used "Motivate Me." In total, 11 health care professionals participated in the focus groups.

Data from the issue log are summarized in Table 1, with relatively few issues with the apps. We asked patients and health care professionals whether they used the "How to guides."

Patients and health care professionals did not feel that they needed to use the guides.

Data were summarized under the two different smartphone apps and then the themes, barriers, facilitators, and building functionality, with 9 further subthemes. There were 2 additional cross-cutting themes: phone usability and confidence. First, we present the cross-cutting themes and then individual themes (Multimedia Appendix 5). Some themes only occurred in health care professional or patient data.

Table 1. Issue log data.

Participant and issue	Number of occurrences	Changes made to app
Patients		
Delay in sync between the health care professional's phone and patient's phone	1	Fast sync capability added
Patients missing messages	2	Pop-up stays on phone until patient says "ok" or "I like it"
Health care professionals		
Amending exercises time consuming on health care professional's phone	1	Process streamlined
Setting goals because of new year. 2017 options not available	1	Glitch in system amended immediately

Cross-Cutting Themes

Phone Usability

Two patients found phone usability difficult in relation to the touch screen. Both had arthritis in their fingers and difficulty using the phone, and one patient was still able to use the phone with practice. The other patient decided that she did not like smartphones and stopped using it altogether. Others had no difficulties using it, even those who had never used a smartphone before. One patient wanted to keep the phone at the end of the study and asked the team further details about purchasing one. Two patients already had a Samsung phone and liked the familiarity of the phone. Patients and health care professionals commented on the standard messages sent by the network provider being annoying.

Confidence in Technology

Having confidence to use different technologies was an important factor in whether they were adopted for both health care professionals and patients. Health care professionals had a fear that the technology would put some patients off their rehabilitation program. Not all of the health care professionals supporting the usability testing were very technology minded and thus felt that there was a learning curve. However, it was observed that showing confidence was important for patients for them to have confidence. There was a view from health care professionals and patients that, as older adults used phones generally in life, confidence would increase.

For patients, there was confidence to be able to use the technology in the first place, and some patients regardless of support did not really engage with the technology. All of the patients who tried using the phone increased their confidence and ability. Once patients had built their confidence, they could then utilize the smartphone for other purposes.

Family support was important in building patients' confidence, whether this was by children or spouses. Health care professionals reported that most families would support patients to use the technology.

Barriers

Types and Delivery of Messages

Patients at times missed motivational messages and prompts in "My Activity programme." Messages were delivered as pop-ups in the app, with no sound given, and they stayed on the screen for 5 min. For those with an existing phone, this was particularly the case because they were given a study phone rather than using the apps on their own phones. Patients received a daily message in the app asking about their health and giving them the option to suspend the messages if they were unwell, the format of this message caused particular confusion for one participant.

We asked patients about the potential of receiving voice messages rather than pop-up messages and whether they would find them more motivating. In general, patients felt that it could be intrusive. The family member of a patient said that if a health care professional was going to send voice messages, prerecorded or in real-time, they may as well ring them.

Health care professionals had mixed views about the potential for sending voice messages, and there was a concern that patients would not want to hear their voices. Health care professionals again spoke about the potential for these kinds of messages to be intrusive, unless in certain contexts. In the context of patients' exercises, they had to be motivational as they felt that instructional messages related to exercises could cause risk if the health care professional could not see what they were doing.

Icons and Pictures

The pictures of the exercises including the use of a chair in "Motivate Me" and "My Activity Programme" were raised as an issue by health care professionals in all 3 services. This was also an issue in the home exercise booklets they handed out. Patients would follow the pictures rather than follow the therapists' direct instructions. Health care professionals encourage patients to carry out their exercises in the kitchen using the worktop for support, rather than a chair. Patients did not comment on the icons or pictures.

App Flow

There were few changes suggested by participants for "My Activity Programme." However, to enable the app to be simplified, there were suggested changes for "Motivate Me." Health care professionals felt that the app was not easy enough to use and could be made more intuitive, particularly when setting goals with patients. There was an issue with the app fitting in their way of practice. Some services first ask patients what they want to achieve and pick the exercises related to those outcome-based goals. However, the team we worked with said that because it was an evidence-based program, they deliver all of the exercises as long as appropriate and tailor the feedback they gave to the patient dependent on their goal.

There was a suggestion that setting both types of goals (outcome and behavioral) in one process led to errors and a risk of work loss if a mistake was made. They suggested that the process needed to be broken up into 2 stages.

Facilitators

The Apps as a Communication Tool

My Activity Programme

Health care professionals were initially worried about us asking patients to report individual exercises, suggesting that it might be too demanding. We decided to test how much they would be willing to report with the option of reducing this if feedback was negative. Patients were happy to report their exercises and found it a satisfying way of communicating with their health care professional. They also liked to receive messages and found them helpful, although some would have liked to be able to respond to messages.

Motivate Me

Health care professionals saw their app as a good way to communicate with patients and see what they had reported. If the patient had not reported their exercises, then the health care professional could use the app to provide encouragement. They could also use it as a communication tool to support patients when they are not with them. It was not only seen as a communication tool in the short term during regular contact but also for long-term follow-up of patients after discharge.

Good App Usability

Patients reported good usability for "My Activity Programme" telling health care professionals during the testing that it was easy to use. There were very few suggested changes. Patients also reported that they could view their exercises on the app with no issues, even those who did not like using the smartphone managed to navigate the app.

Goal-Setting Functions

Goal setting is an important part of the behavioral intervention within "Motivate Me." A part of usability was that health care professionals felt that it fit within their existing practice. Patients also liked the idea of goal setting together with the health care professional using the app, feeling that their needs and expectations were being considered.

Flexibility of Use

Health care professionals discussed the flexibility of the smartphone apps and how they could be used with different populations and services. They thought they could be used as preventative apps for exercise instructors and charitable organizations who may be delivering evidence-based fall prevention exercises. They also felt that they could be important tools for other at-risk populations.

Building Functionality

More Flexibility in Times

In terms of the functions in "Motivate Me," there were some changes that patients asked for when setting goals. They did not always want to specify a time when they were going to do home exercise; instead, they wanted to fit it around their daily lives. We asked them whether they were happy to schedule a day to exercise, and they preferred to set specific days without specific times. Patients were happy to schedule a time when they were exercising with the health care professional, so flexibility was required within the app.

Additional Information

Motivate Me

Health care professionals thought it would be helpful for patients if they could add notes underneath where they had set the exercises, as they did in home exercise booklets. They discussed how it would be useful to use a keyword to search for goals within the library.

My Activity Programme

Health care professionals and patients suggested that patients should be able to view their goals (and exercises) at any point. In the prototype, patients could only view them on the days they had set to exercise, and it was felt that this caused inflexibility and confusion. Health care professionals wanted to check the patients' phone so they could see that the goals had gone into their app, in the current format they could not do this.

Patients and health care professionals talked about whether the motivational messages could stay on the screen for longer so they did not get missed. Some patients requested a loud noise when the pop-up came through to prevent them from missing it. Health care professionals also talked about not knowing whether patients had seen the messages and whether we could ask patients to say they "liked them," and then see this in their app.

Finally, one service discussed whether "My Activity Programme" could include videos of the exercises. There was a discussion about patients sending health care professionals' video clips of them carrying out the exercises. However, if this was too complex, then health care professionals discussed the potential to at least link the app to other resources, existing systems, and websites they already had in place where patients could access videos.

Discussion

Principal Findings

Introducing technology to patients and health care professionals is challenging, with a range of barriers around usability and acceptability. However, although barriers to the use of the apps were identified within this study, solutions were also offered by patients and health care professionals. Getting users involved in the development of technologies is recommended as key to their success, and we think this has led to very few changes suggested to the "My Activity Programme" during the implementation phase [21]. The iterative approach used through HCD has been found to be successful in creating apps that are usable for older adults and clinicians [34,36] and has enabled us to ensure that the apps are flexible and can continue to evolve. Our app development fits within a user-led design approach and we have consulted older adults, patients, and health care professionals not only about their needs and at key design points but also throughout the design process [49]. We liaised with different sets of health care professionals and both older adults in the community, previous patients of fall services, and current patients through the different stages of development to enable diverse inputs. Overall, we found that creating two apps and making "My Activity Programme" simple, tailored, and personalized helped to make it more acceptable to patients and alleviated fears of technology.

Results from the analysis phase indicated that goal setting (outcomes and behavior) and feedback would enhance current practice. Collecting example goals from both health care professionals, instructors, and older adults was an important part of ensuring that the apps could be personalized, which is essential [50]. Using recognizable and relevant goals and icons for the exercises were thought to make professionals and patients feel more confident.

The results of the implementation phase reflected earlier findings from the analysis phase, suggesting that there were initial barriers to phone use. Issues with the touch screen were discussed as a barrier to use, particularly where patients had arthritis and poor finger dexterity. The TAM focuses on perceived ease of use and the role it plays in acceptance [37]. A smartphone is needed for these types of apps, and from previous research, we know that patients would prefer not to navigate new technology [14]. All current smartphones are provided with a touch screen, and further work is being carried out to explore their use [51]. However, it was suggested that the use of a smartphone pen would overcome most of the issues. Text messages from network providers were perceived by health care professionals as annoying, the option to use roaming SIM cards (across networks) to increase 3G/4G coverage now means that this does not have to be an issue.

Ensuring that the apps fit the needs of health care professionals and patients was an important part of our development process. Confidence in using the technology is important for health care professionals and patients. Health care professionals are more likely to adopt technological interventions if they perceive the benefit to themselves and patients [52]. The TAM focuses on the acceptance of technology being related to its perceived usefulness [37]. Health care professionals saw their confidence as important in encouraging patient confidence in the use of the technology. There was a perception from health care professionals and patients that the more they used the technology the more confident they would become, something established in previous research [52,53]. The support of family and friends played an important role in supporting patients' attitudes toward technology and confidence to use it [21].

Feedback on the apps provides further insight into the potentially acceptable delivery of messages. We found that pop-up messages were mostly acceptable and a key communication tool between health care professionals and patients. Personalized and tailored messages directly from the health care professional garnered the best response in our testing, and a recent evidence review supports this [54]. There is currently no clear evidence on what older adults find motivating in terms of number, type, and delivery of messages from an app that supports fall rehabilitation. We know that to encourage participation in fall prevention technologies, we should emphasize the benefits of positive active aging [21] and that messages should be positively framed [44,55].

For the apps to be acceptable, they had to fit with health care professionals' practice as well as be flexible and fit with older adults' needs. Previous research suggests that technological solutions for fall prevention need to be individually tailored [21]. Interviews with patients suggested flexibility around goal setting and message delivery so that the use of the technology was not restrictive and fit with their lives. Focus groups with staff-provided suggestions of how "Motivate Me" could be improved so that further individualization was offered, for example, adding notes and ensuring the process of goal setting more closely reflected practice. Health care professionals discussed other additional features that could enhance the app, for example, videos, which need to be carefully considered alongside burden on the participant and data confidentiality, issues previously raised as barriers [21]. Health care professionals also identified the potential for the apps to be used with different populations and for follow-up after the rehabilitation phase.

"My Activity Programme" included asking patients to self-report their rehabilitation exercises. There is a lack of well-validated self-report measures for recording adherence in the specific context of prescribed but unsupervised home-based rehabilitation exercises for older adults [56]. Studies have found a poor correlation between self-reported exercise questionnaires and objective measures such as accelerometers [57,58]. However, Fukuoka et al [59] found good compliance and correlation between self-reported activity on an app and a step counter. Objective measurement of rehabilitation exercises is difficult without asking patients to wear multiple sensors [60], something that has shown promise but has usability issues, particularly with this patient cohort [21].

During the development phase, health care professionals questioned whether participants would be willing to report individual exercises, although health care professionals found information on type of exercise, dose, and intensity useful. We found that patients were happy to report their exercises and, at

times, found this motivation, which is related to the general literature on self-monitoring [61]. This app could in fact provide important contextual data for those trying to monitor people's physical activity as the sensors in the phone alone are currently inadequate at detecting static strength and balance [60].

Limitations

There were limitations to the study. As part of the analysis phase of the study, we decided to move to a system that used 2 apps that interacted (setting goals on "Motivate Me" and using "My Activity Programme" to receive messages and report exercises). This approach mirrored practices as the health care professional will set the patients' exercise program in consultation with them. It also made the patient app easier to use. We did not ask older adults in the workshop or patients in the interviews directly about whether they would like to be able to change their exercise schedule or goals themselves, but instead asked open-ended questions about the apps. This never arose as an issue for either older adults in the workshop or patients in interviews. However, it may be that some patients could find that the inability to change their program themselves (eg, when they planned to exercise and their outcome goal) negatively impacts their motivation; we know that control is important to use [21]. We need to balance functionality with the capabilities of our target group, and this is an important consideration for future iterations and something that may occur with longer use. There were very few changes to "My Activity Programme" suggested by older adults in the workshop. This could be because at this initial stage, not all of the functionality was in place (eg, messages did not come through at this point); therefore, they mostly focused on its look and style. Apart from explaining how the goal setting would work with health care professionals and asking for feedback on ideas for goals and messages, we did not show older adults "Motivate Me" in the first workshop. This was because we primarily wanted feedback from health care professionals at this stage and did not want to overwhelm older adults with something they would not directly use. However, further data from older adults could have proved useful.

The workshop was carried out with community-dwelling older adults who attended a strength and balance class, whereas the testing was carried out directly with patients. It could be argued that they were two different populations, and this may have influenced feedback. We argue that it is a strength to represent a wide variety of older adults' views. We have a good representation of men and women in terms of older adults, patients, and health care professionals. Although usability testing was performed with patients aged 50 years and older, we wanted to test the app directly with the population who would use it. Recruitment took longer than anticipated and was exacerbated by the Christmas period. Therefore, a much smaller number of patients were recruited than planned. There was a lack of recruitment from Black and minority ethnic populations. However, the participants recruited represented a good mix of patients in terms of comorbidities, age, gender, and prior experience with technology. The period we tested the technology over was relatively short and may not have identified all implementation and usability issues. We also could have used a usability questionnaire. However, at this point, this was a predominantly qualitative initial development, acceptability, and usability work designed to lead to further feasibility work.

Conclusions

Overall, we have established that "Motivate Me" and "My Activity Programme" were acceptable to most older adults and patients who participated and all health care professionals. We developed important personalized content for our apps. There is a lack of research on smartphone-based interventions for the support of fall management and prevention. This study enabled us to build 2 apps that have the potential to support health care professionals and patients with their rehabilitation program. "Motivate Me" and "My Activity Programme" require further improvement and development and then need to be explored in practice as part of a bigger feasibility trial.

Acknowledgments

The authors thank Professor Dawn Skelton and LLT for allowing them to use their images and the name "Motivate Me" for the health professional app. The authors also thank 2 PPI representatives who sit on their advisory group. This study is an independent research arising from an National Institute for Health Research Postdoctoral Fellowship Award, Dr Helen Hawley-Hague, (PDF -2015-08-012) supported by the National Institute for Health Research**.** 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.

Authors' Contributions

HH led the research project and its design, managed the study overall, led the development of the apps, and led the writing of the manuscript. CT and SM provided technical support, created the apps for the study, and advised on outcomes and the manuscript. JH, LC, CT, and SM have provided scientific advice around the design of the study and commented on the manuscript. EM led the PPI work and gave advice on the operationalization of the study and commented on the manuscript. CF assisted with the coding and analysis and contributed to the manuscript.

Conflicts of Interest

LC, SM, and CT own a share of the University of Bologna spin-off company, mHealth Technologies srl. CT is the president of mHealth Technologies srl.

Multimedia Appendix 1
Functions of the applications at each stage of development.
[DOCX File , 13 KB - mhealth_v8i9e15460_app1.docx]

Multimedia Appendix 2
Interview and focus group schedules.
[DOCX File , 15 KB - mhealth_v8i9e15460_app2.docx]

Multimedia Appendix 3
Demographics of participants for workshops and usability study.
[DOCX File , 14 KB - mhealth_v8i9e15460_app3.docx]

Multimedia Appendix 4
Feedback from workshops and corresponding changes made.
[DOCX File , 16 KB - mhealth_v8i9e15460_app4.docx]

Multimedia Appendix 5
Themes from interviews and focus groups.
[DOCX File , 21 KB - mhealth_v8i9e15460_app5.docx]

References

1. Falls and Fractures: Consensus Statement and Resources Pack. Government of UK. 2017. URL: https://www.gov.uk/government/publications/falls-and-fractures-consensus-statement [accessed 2019-05-29] [WebCite Cache ID 78jlzoo8S]
2. Gillespie L, Robertson M, Gillespie W, Lamb SE, Gates S, Cumming RG, et al. Interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2009 Apr 15(2):CD007146. [doi: 10.1002/14651858.CD007146.pub2] [Medline: 19370674]
3. Sherrington C, Michaleff ZA, Fairhall N, Paul SS, Tiedemann A, Whitney J, et al. Exercise to prevent falls in older adults: an updated systematic review and meta-analysis. Br J Sports Med 2017 Dec;51(24):1750-1758. [doi: 10.1136/bjsports-2016-096547] [Medline: 27707740]
4. El-Khoury F, Cassou B, Charles M, Dargent-Molina P. The effect of fall prevention exercise programmes on fall induced injuries in community dwelling older adults. Br J Sports Med 2015 Oct;49(20):1348. [doi: 10.1136/bmj.f6234] [Medline: 26429908]
5. Public Health England. Muscle and Bone Strengthening and Balance Activities for General Health Benefits in Adults and Older Adults. Government of UK Developer Documentation. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/721874/MBSBA_evidence_review.pdf [accessed 2020-09-04]
6. Benzinger P, Becker C, Todd C, Bleibler F, Rothenbacher D, König HH, et al. The impact of preventive measures on the burden of femoral fractures - a modelling approach to estimating the impact of fall prevention exercises and oral bisphosphonate treatment for the years 2014 and 2025. BMC Geriatr 2016 Apr 01;16:75 [FREE Full text] [doi: 10.1186/s12877-016-0247-9] [Medline: 27038629]
7. Robertson MC, Gardner MM, Devlin N, McGee R, Campbell AJ. Effectiveness and economic evaluation of a nurse delivered home exercise programme to prevent falls. 2: Controlled trial in multiple centres. BMJ 2001 Mar 24;322(7288):701-704 [FREE Full text] [doi: 10.1136/bmj.322.7288.701] [Medline: 11264207]
8. Skelton D, Dinan S, Campbell M, Rutherford O. Tailored group exercise (Falls Management Exercise -- FaME) reduces falls in community-dwelling older frequent fallers (an RCT). Age Ageing 2005 Nov;34(6):636-639. [doi: 10.1093/ageing/afi174] [Medline: 16267192]
9. Older People's Experience of Therapeutic Exercise as Part of a Falls Prevention Service. RCP London. 2012. URL: https://www.rcplondon.ac.uk/projects/outputs/older-peoples-experience-therapeutic-exercise-part-falls-prevention-service [accessed 2020-09-16]
10. Davis JC, Robertson MC, Ashe MC, Liu-Ambrose T, Khan KM, Marra CA. Does a home-based strength and balance programme in people aged > or =80 years provide the best value for money to prevent falls? A systematic review of economic evaluations of falls prevention interventions. Br J Sports Med 2010 Feb;44(2):80-89. [doi: 10.1136/bjsm.2008.060988] [Medline: 20154094]
11. A Return on Investment Tool for the Assessment of Falls Prevention Programmes for Older People Living in the Community. Developer Documentation. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/679856/A_return_on_investment_tool_for_falls_prevention_programmes.pdf [accessed 2020-09-16]
12. Carpenter H, Audsley S, Coupland C, Gladman J, Kendrick D, Lafond N, et al. PHysical activity Implementation Study In Community-dwelling AduLts (PHISICAL): study protocol. Inj Prev 2019 Oct;25(5):453-458. [doi: 10.1136/injuryprev-2017-042627] [Medline: 29305397]

13. Hawley H. Older adults' perspectives on home exercise after falls rehabilitation: Understanding the importance of promoting healthy, active ageing. Health Education Journal 2009 Sep 29;68(3):207-218. [doi: 10.1177/0017896909339533]
14. Fukuoka Y, Kamitani E, Bonnet K, Lindgren T. Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: a focus group analysis. J Med Internet Res 2011 Jul 14;13(3):e49 [FREE Full text] [doi: 10.2196/jmir.1770] [Medline: 21752785]
15. Vankipuram M, McMahon S, Fleury J. ReadySteady: app for accelerometer-based activity monitoring and wellness-motivation feedback system for older adults. AMIA Annu Symp Proc 2012;2012:931-939 [FREE Full text] [Medline: 23304368]
16. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
17. Worringham C, Rojek A, Stewart I. Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. PLoS One 2011 Feb 09;6(2):e14669 [FREE Full text] [doi: 10.1371/journal.pone.0014669] [Medline: 21347403]
18. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
19. Boulos M, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
20. Rise of the Social Seniors Revealed. OfCom. 2017. URL: https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/rise-social-seniors [accessed 2020-09-04]
21. Hawley-Hague H, Boulton E, Hall A, Pfeiffer K, Todd C. Older adults' perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. Int J Med Inform 2014 Jun;83(6):416-426. [doi: 10.1016/j.ijmedinf.2014.03.002] [Medline: 24798946]
22. Rasche P, Mertens A, Bröhl C, Theis S, Seinsch T, Wille M, et al. The "Aachen fall prevention App" - a Smartphone application app for the self-assessment of elderly patients at risk for ground level falls. Patient Saf Surg 2017;11:14 [FREE Full text] [doi: 10.1186/s13037-017-0130-4] [Medline: 28503199]
23. Delbaere K, Valenzuela T, Woodbury A, Davies T, Yeong J, Steffens D, et al. Evaluating the effectiveness of a home-based exercise programme delivered through a tablet computer for preventing falls in older community-dwelling people over 2 years: study protocol for the Standing Tall randomised controlled trial. BMJ Open 2015 Oct 22;5(10):e009173 [FREE Full text] [doi: 10.1136/bmjopen-2015-009173] [Medline: 26493461]
24. Sandlund M, Lindgren H, Pohl P, Melander-Wikman A, Bergvall-Kåreborn B, Lundin-Olsson L. Towards a Mobile Exercise Application to Prevent Falls: A Participatory Design Process. Semantic Scholar. URL: https://pdfs.semanticscholar.org/177d/66a2ab9877c07a73ba20715b71bedaf90152.pdf [accessed 2020-09-04]
25. Hawley-Hague H, Horne M, Campbell M, Demack S, Skelton DA, Todd C. Multiple levels of influence on older adults' attendance and adherence to community exercise classes. Gerontologist 2014 Aug;54(4):599-610 [FREE Full text] [doi: 10.1093/geront/gnt075] [Medline: 23899623]
26. Yardley, Donovan-Hall M, Francis K, Todd C. Attitudes and beliefs that predict older people's intention to undertake strength and balance training. J Gerontol B Psychol Sci Soc Sci 2007 Mar;62(2):P119-P125. [doi: 10.1093/geronb/62.2.p119] [Medline: 17379672]
27. Ajzen I. Attitudes, Personality and Behaviour. In: Attitudes, Personality and Behaviour. Maidenhead, UK: Open University Press; 1988.
28. Chatzisarantis N, Hagger M. Exercise psychology effects of a brief intervention based on the theory of planned behavior on leisure-time physical activity participation. J Sport Exerc Psychol 2005;27:470-487 [FREE Full text] [doi: 10.1123/jsep.27.4.470]
29. Epton T, Currie S, Armitage C. Unique effects of setting goals on behavior change: Systematic review and meta-analysis. J Consult Clin Psychol 2017 Dec;85(12):1182-1198. [doi: 10.1037/ccp0000260] [Medline: 29189034]
30. McEwan D, Harden S, Zumbo B, Sylvester B, Kaulius M, Ruissen G, et al. The effectiveness of multi-component goal setting interventions for changing physical activity behaviour: a systematic review and meta-analysis. Health Psychol Rev 2016;10(1):67-88. [doi: 10.1080/17437199.2015.1104258] [Medline: 26445201]
31. Koestner R, Hope N. A self-determination theory approach to goals. In: Gagné M, editor. The Oxford Handbook of Work Engagement, Motivation, and Self-Determination Theory. Oxford, UK: Oxford University Press; 2014.
32. Hawley-Hague H, Horne M, Skelton D, Todd C. Older Adults' Uptake and Adherence to Exercise Classes: Instructors' Perspectives. J Aging Phys Act 2016 Jan;24(1):119-128. [doi: 10.1123/japa.2014-0108] [Medline: 26214265]
33. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
34. Rosenzweig E. Successful User Experience: Strategy and Roadmaps. New York, USA: Elsevier Inc; 2015.

35. Sekhon M, Cartwright M, Francis J. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
36. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-Centered Design Study: Enhancing the Usability of a Mobile Phone App in an Integrated Falls Risk Detection System for Use by Older Adult Users. JMIR Mhealth Uhealth 2017 May 30;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]
37. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
38. Venkatesh V, Morris M, Davis G, Davis F. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425 [FREE Full text] [doi: 10.2307/30036540]
39. Developing and Evaluating Complex Interventions: Following Considerable Development in the Field Since 2006, MRC and NIHR Have Jointly Commissioned an Update of This Guidance to Be Published in 2019. Medical Research Council. 2008. URL: http://www.mrc.ac.uk/complexinterventionsguidance [accessed 2020-09-04]
40. Larman C, Basili V. Iterative and Incremental Developments. A Brief History. New York, USA: A Brief History; Jun 2003.
41. Later Life Training. URL: https://www.laterlifetraining.co.uk/ [accessed 2020-03-01]
42. Hawley-Hague H, Tacconi C, Mellone S, Martinez E, Easdon A, Yang FB, et al. Can smartphone technology be used to support an effective home exercise intervention to prevent falls amongst community dwelling older adults?: the TOGETHER feasibility RCT study protocol. BMJ Open 2019 Sep 18;9(9):e028100 [FREE Full text] [doi: 10.1136/bmjopen-2018-028100] [Medline: 31537557]
43. Johnson R, Onwuegbuzie AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. New York, USA: Educational Researcher; Jul 2016.
44. Preliminary Guidelines for Design and Implementation of Technologies. FARSEEING. 2014. URL: http://farseeingresearch.eu/wp-content/uploads/2014/07/Preliminary-Guidelines-V1.0-Public.pdf [accessed 2020-09-04]
45. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Hawley-Hague H, Roden A, Abbott J. The evaluation of a strength and balance exercise program for falls prevention in community primary care. Physiother Theory Pract 2017 Aug;33(8):611-621. [doi: 10.1080/09593985.2017.1328721] [Medline: 28590815]
48. del Rosario M, Redmond S, Lovell N. Tracking the Evolution of Smartphone Sensing for Monitoring Human Movement. Sensors (Basel) 2015 Jul 31;15(8):18901-18933 [FREE Full text] [doi: 10.3390/s150818901] [Medline: 26263998]
49. Schuler D, Namioka A. Participatory Design: Principles and Practices. Boca Raton, FL: CRC Press; 2009.
50. Helbostad JL, Vereijken B, Becker C, Todd C, Taraldsen K, Pijnappels M, et al. Mobile Health Applications to Promote Active and Healthy Ageing. Sensors (Basel) 2017 Mar 18;17(3). [doi: 10.3390/s17030622] [Medline: 28335475]
51. Danial-Saad A, Chiari L. A multidisciplinary approach for developing an assessment tool for touch screen devices. Disabil Rehabil Assist Technol 2018 Nov;13(8):745-753. [doi: 10.1080/17483107.2017.1370500] [Medline: 28903629]
52. MacNeill V, Sanders C, Fitzpatrick R, Hendy J, Barlow J, Knapp M, et al. Experiences of front-line health professionals in the delivery of telehealth: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e401-e407 [FREE Full text] [doi: 10.3399/bjgp14X680485] [Medline: 24982492]
53. Vaportzis E, Clausen MG, Gow AJ. Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. Front Psychol 2017 Oct 04;8:1687. [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
54. Sahin C, Courtney KL, Naylor PJ, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: A systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]
55. Notthoff N, Klomp P, Doerwald F, Scheibe S. Positive messages enhance older adults' motivation and recognition memory for physical activity programmes. Eur J Ageing 2016 Sep;13(3):251-257 [FREE Full text] [doi: 10.1007/s10433-016-0368-1] [Medline: 28804382]
56. Bollen JC, Dean SG, Siegert RJ, Howe TE, Goodwin VA. A systematic review of measures of self-reported adherence to unsupervised home-based rehabilitation exercise programmes, and their psychometric properties. BMJ Open 2014 Jun 27;4(6):e005044. [doi: 10.1136/bmjopen-2014-005044] [Medline: 24972606]
57. Falck RS, McDonald SM, Beets MW, Brazendale K, Liu-Ambrose T. Measurement of physical activity in older adult interventions: a systematic review. Br J Sports Med 2016 Apr;50(8):464-470. [doi: 10.1136/bjsports-2014-094413] [Medline: 26276362]
58. Prince S, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]

59. Fukuoka Y, Kamitani E, Dracup K, Jong S. New insights into compliance with a mobile phone diary and pedometer use in sedentary women. J Phys Act Health 2011 Mar;8(3):398-403 [FREE Full text] [doi: 10.1123/jpah.8.3.398] [Medline: 21487139]
60. Correia F, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Home-based rehabilitation with a novel digital biofeedback system versus conventional in-person rehabilitation after total knee replacement: a feasibility study. Sci Rep 2018 Jul 26;8(1):11299 [FREE Full text] [doi: 10.1038/s41598-018-29668-0] [Medline: 30050087]
61. Helsel DL, Jakicic JM, Otto AD. Comparison of techniques for self-monitoring eating and exercise behaviors on weight loss in a correspondence-based intervention. J Am Diet Assoc 2007 Oct;107(10):1807-1810. [doi: 10.1016/j.jada.2007.07.014] [Medline: 17904942]

Abbreviations

BCT: behavior change technique
FaME: falls management exercise
HCD: human-centered design
LLT: later life training
PPI: patient and public involvement
TAM: technology acceptance model
TPB: theory of planned behavior
SDT: self-determination theory

Edited by G Eysenbach; submitted 11.07.19; peer-reviewed by P Rasche, M Ehn, OY Lo; comments to author 05.01.20; revised version received 01.04.20; accepted 16.06.20; published 28.09.20.

Please cite as:
Hawley-Hague H, Tacconi C, Mellone S, Martinez E, Ford C, Chiari L, Helbostad J, Todd C
Smartphone Apps to Support Falls Rehabilitation Exercise: App Development and Usability and Acceptability Study
JMIR Mhealth Uhealth 2020;8(9):e15460
URL: http://mhealth.jmir.org/2020/9/e15460/
doi:10.2196/15460
PMID:32985992

Original Paper

Gaps in Team Communication About Service Statistics Among Health Extension Workers in Ethiopia: Secondary Data Analysis

Seohyun Lee[1], PhD; Eunji Kim[1], BA; Tekaligne Birhane Desta[1], MA

Yonsei University, Wonju, Republic of Korea

Corresponding Author:
Seohyun Lee, PhD
Yonsei University
1 Yonseidae-gil
Wonju,
Republic of Korea
Phone: 82 33 760 2357
Email: leeesh@yonsei.ac.kr

Abstract

Background: In Ethiopia, health extension workers (HEWs) are deployed across the country by the government to meet public health needs. Team communication is important for effective teamwork, but community health workers in low-resource settings like Ethiopia may face challenges in carrying out team meetings to compile service statistics. This is due to the nature of their outreach activities, which requires extensive travel.

Objective: This study aimed to identify gaps in team communication about service statistics among HEWs in Ethiopia. Considering mobile communication and data collection as tools for bridging these gaps, we examined disparities in access to electricity, which has been identified as one of the major barriers to this approach.

Methods: Data from the most recent Performance Monitoring and Accountability 2020 service delivery point survey were used for our analysis. Logistic regression analysis was performed to identify disparities in team communication on service statistics for family planning, which is a major component of the HEW's job. Disparities were examined across health facilities with different levels of HEW integration in their staffing structure (ie, no HEWs, at least one HEW, or only HEWs). Additionally, a chi-square test was conducted to examine disparities in access to electricity to explore the potential of mobile communication and data collection integration.

Results: In total, 427 health facilities of four different types (ie, hospitals, health centers, health posts, and health clinics) were included in our analysis. At most health posts (84/95, 88%), only HEWs were employed; none of the health clinics integrated the HEW model into their staffing structure. Among the 84 health posts, the odds of having team meetings on family planning service statistics in the past 12 months were 0.48 times the odds of those without HEWs (P=.02). No statistically significant differences were found between HEW-only facilities and facilities with at least one HEW. Most health facilities (69/83, 83.13%) with HEWs as the only staff had no electricity at the time of the survey while 71.25% (57/80) had intermittent access (ie, service disruption lasting 2 or more hours that day). There were statistically significant differences in electricity access among health facilities with different levels of HEW integration (P<.001).

Conclusions: Facilities employing only HEWs were less likely to have regular team meetings to discuss service statistics. Since their responsibilities involve extensive outreach activities, travel, and paper-based recordkeeping, empowering HEWs with mobile communication and data collection can be a workable solution. The empirical evidence regarding disparities in electricity access also supports the need for and the feasibility of this approach.

(JMIR Mhealth Uhealth 2020;8(9):e20848) doi:10.2196/20848

KEYWORDS

team communication; health extension workers; mobile communication; mobile data collection; Ethiopia; health worker; communication; data

Introduction

Team-based interactions among health workers have been considered to be drivers for productivity, effective service delivery, cooperation, and quality improvement [1,2]. When teams communicate on a regular basis, health workers can share workload, make more informed decisions, and exchange new ideas in a more efficient manner [3]. In resource-limited settings, building teamwork through communication can be challenging but critical. In fact, previous studies have demonstrated that communication is one of the key factors for establishing teamwork in low- and middle-income countries (LMICs). A study conducted in Tanzania identified team communication as one of the main strategies for team implementation efforts in improving postpartum care [4]. Another study that reviewed 47 studies regarding teamwork among community health workers (CHWs) and health care teams found regular team meetings with specific purposes to be an essential component of the team's interdependence [2]. Another study from India revealed that team communication for cross-checking records, planning outreach sessions, or cross-training is an important part of practices managed by CHWs [5]. In the Zambian context, Yeboah-Antwi et al [3] identified good communication and sharing information as two core factors that characterize teamwork.

In LMICs where a shortage of skilled health professionals triggered the deployment of CHWs, team communication can be even more challenging. Specifically, scheduling a regular team meeting may not be feasible since the majority of the CHW's work involves community outreach that requires extensive travel. CHWs are in a unique position where they function as a liaison between the community and the health system. According to the World Health Organization (WHO) guidelines, the tasks of CHWs defined by International Labour Organization include providing education and advice for communities, conducting outreach efforts, and managing medical supplies [6]. In other words, CHWs are responsible for supporting communities as well as the health system, requiring frequent travel and activities outside of health facilities.

Ethiopia is one of the low-income countries that implemented a community-based health program involving CHWs called health extension workers (HEWs). In 2003, the Ethiopian Federal Ministry of Health introduced the Health Extension Program, and as part of that program, HEWs have been recruited, trained, and salaried by the government to meet the public health needs of the community [7-9]. HEWs are mostly female with at least a 10th-grade education and are from the communities they serve, allowing them to build trust and rapport with the community [8]. According to task analysis studies for HEWs, they spend a majority of their time on travel or outreach activities, which may limit the possibilities for effective team communication. Studies have noted that HEWs spend approximately 75% of their working time performing outreach activities [8,10,11]. A time and motion study found that they spend 15.5% of their time traveling between work activities [12]. As a result, gathering for a regular staff meeting at the same location may not be a workable solution for HEWs to share information or ideas.

Indeed, the HEW's difficulty due to extensive travel lies not only in the flow of information per se but also in the quality of information shared among HEWs and their supervisors [13]. Previous literature suggested that the existing paper-based data collected between travels by HEWs pose a challenge in terms of poor quality and extra burden for aggregation or reporting [14,15]. Studies have pointed out that recordkeeping and reporting of service data is one of the main tasks assigned to HEWs, requiring them to bring a paper-based field book while traveling for outreach activities [14,16]. After returning to health facilities, HEWs need to spend extra time on compiling or validating the data and generating service statistics for reporting. Working with paper-based field book can be inefficient, and there is a risk for loss of data or limited validity.

In this context, the primary objective of this study is to identify gaps in team communication on service statistics among health facilities with no HEWs, at least one HEW, or exclusively HEWs comprising their staffing structure. Based on previous literature on the impact of mobile technology on teamwork and communication, we considered mobile communication and data collection as good candidates for bridging these gaps [17]. To identify potential challenges in mobile communication and data collection, we examined disparities in access to electricity across different health facilities. In fact, evidence from a systematic review on health workers' use of mobile devices in low-resource settings found that poor access to electricity is one of the main barriers to mobile communication and data collection [18].

Methods

Data and Scope

For analysis, the most recent Performance Monitoring and Accountability 2020 (PMA2020) service delivery point data were used. PMA2020 is an initiative for improving family planning in LMICs supported by the Bill and Melinda Gates Foundation and collects a nationally representative sample data through mobile data collection [19]. Specifically, the PMA2020 Ethiopia Round 6 Service Delivery Point 2018 data were used for the study. PMA2020 service delivery point data comprise a nationally representative data set collected at the health facility level and is publicly available upon request [19]. For data collection, trained resident enumerators visit the sampled facilities and conduct surveys via mobile phones [20]. By utilizing mobile data collection strategy for large-scale surveys on various health issues such as family planning or water and sanitation, PMA2020 provides rapid and high-quality data for decision making and policy design.

Ethiopia Round 6 data contain information about 476 health facilities and their service provision and quality, with special attention paid to family planning. In total, the 427 health facilities that reported valid responses to the key questions needed for our analysis were included in this study. For example, facilities that did not provide information about the number of each staff position were excluded since disparities were compared across health facilities with different levels of HEW integration in the staffing structure. Health facilities were categorized as those with no HEWs, at least one HEW, or HEWs as the only staff position. In the context of Ethiopia's health

system, two HEWs are usually deployed to each health post to cover 3000 to 5000 people to improve primary health care [16]. In addition to health posts, three other types of health facilities were included for analysis—hospitals, health centers, and health clinics. Although data on pharmacies, retail outlets, and others were available, they were not included in the analysis because their focus is not on the provision of health services.

Measurements

The primary outcome measure is the binary variable of team communication on service statistics for family planning, which is the main responsibility of the HEW [9,16,21]. It is defined as having any meetings with staff to discuss service statistics or inventory for family planning in the past 12 months [19]. Examples of family planning service statistics include the total number of visits or the number of emergency contraception units sold in the past completed month [19]. Explanatory variables include the level of integration of HEWs into the staffing structure, location (ie, urban or rural areas), knowledge of the catchment population size, and recent external supervisor visit and support from nongovernmental organizations (NGOs) for family planning services. The level of integration of HEWs into the staffing structure was measured categorically—facilities with no HEWs, with at least one HEW, or comprising entirely of HEWs. Recent external supervisor visits were categorized as having occurred more than 6 months ago versus within the past 6 months. Support from NGOs for family planning services included funding or other types of support in the past 12 months, like training, technical assistance, or supplies [19]. The reasoning behind adding this variable comes from previous studies that discussed NGOs' special focus on management strategies and relatively higher level of technical competence than government or private facilities for the implementation of family planning services in Ethiopia [22,23].

Having access to electricity can facilitate wireless communication, charging mobile devices, and maintaining other electronic devices for communication and data management. To assess the disparities in access to electricity as the potential barrier to mobile communication and data collection, two measures related to electricity access were examined. The first measure looked at whether the facility had electricity at the time of the survey, and the second measure investigated whether there was intermittent access at the facility (ie, service disruption of 2 or more hours on the day the survey was conducted).

Statistical Analysis

Descriptive statistics were used to provide an overview of the characteristics of health facilities that have no HEWs, at least one HEW, or comprised only HEWs. To study the association between team meetings on service statistics and the integration of HEWs into the staffing structure, we conducted a binary logistic regression analysis. Predicted probabilities for having team communication in the past 12 months among health facilities were estimated and presented graphically. A chi-square test was performed to assess disparities in access to electricity among health facilities with no HEWs, at least one HEW, or only HEWs. Stata (version 15, StataCorp LLC) was used for statistical analysis [24].

Results

Descriptive Statistics

Descriptive statistics are presented in Table 1. In terms of facilities, four types representing the health system of Ethiopia—hospitals, health centers, health posts, and health clinics—were identified. The hierarchy of the primary health care unit in Ethiopia is as follows: the *woreda* (district) health office oversees health centers, which in turn supervise five health posts [16,25]. Additionally, hospitals and health clinics provide support for the health system. In total, 84 out of 95 health posts (88%) comprised only HEWs; no other facility type had this composition. There was at least one HEW deployed at all health posts. On the other hand, health clinics in the sample did not integrate HEWs at all into their staffing structure. Among facilities that had no HEWs, 45.1% (n=105) were hospitals, 45.9% (n=107) were health centers, and 9% (n=21) were health clinics.

Table 1. Descriptive statistics for health facilities (source: PMA2020 Ethiopia Round 6 Service Delivery Point 2018 [19]).

Variable	Staffing structure, n (%)		
	No HEWs[a] (n=233)	At least one HEW (n=110)	Only HEWs (n=84)
Facility type			
Hospital	105 (45.1)	3 (2.7)	0 (0)
Health center	107 (45.9)	96 (87.3)	0 (0)
Health post	0 (0)	11 (10)	84 (100)
Health clinic	21 (9.0)	0 (0)	0 (0)
Location			
Rural	122 (52.4)	41 (37.3)	73 (86.9)
Urban	111 (47.6)	69 (62.7)	11 (13.1)
Managing authority			
Government	210 (90.1)	110 (100)	84 (100)
Private	22 (9.4)	0 (0)	0 (0)
Other	1 (0.4)	0 (0)	0 (0)
Knowledge of catchment population size			
Not applicable (no catchment area)	17 (7.3)	4 (3.6)	1 (1.2)
No	18 (7.7)	5 (4.6)	1 (1.2)
Yes	198 (85.0)	101 (91.8)	82 (97.6)
Recent external supervisor visit			
More than 6 months ago	24 (10.3)	21 (19.1)	6 (7.1)
Within the past 6 months	209 (89.7)	89 (80.9)	78 (92.9)
Received support in the past 12 months from an NGO[b] for family planning services			
No	101 (43.4)	26 (23.6)	50 (59.5)
Yes	132 (56.7)	84 (76.4)	34 (40.5)
Had meetings with staff to discuss services statistics for family planning in the past 12 months			
No	45 (19.3)	12 (10.9)	24 (28.6)
Yes	188 (80.7)	98 (89.1)	60 (71.4)

[a]HEW: health extension worker.

[b]NGO: nongovernmental organization.

With regard to location, 86.9% (n=73) of the facilities with only HEWs were located in rural areas, whereas 62.7% (n=69) of those with at least one HEW were located in urban areas. Among health facilities with no HEWs, 52.4% (n=122) were located in rural areas and 47.6% (n=111) were in urban areas. All facilities with at least one HEW or HEWs as the only staff were managed by the government. This result reflects the fact that HEWs are employed, trained, and salaried by the government [10,26].

When asked about the size of the catchment population, facilities with HEWs as the only position had the highest proportion of positive answers (n=82, 97.6%). For facilities with no HEWs or at least one HEW, more than 5% (n=35 and 9, respectively) answered that they had no catchment area or did not know the size of the catchment population.

In terms of recent external supervisor visits, facilities with HEWs as the only staff had the highest proportion of having a supervisor visit within the past 6 months (n=78, 92.9%). For facilities with no HEWs or at least one HEW, it was 89.7% (n=209) and 80.9% (n=89), respectively.

As for support from NGO for family planning services in the last 12 months, facilities with at least one HEW had the highest proportion of having received this support (n=84,76.4%); 56.7% (n=132) of facilities with no HEWs and 40.5% (n=34) of those with HEWs as the only staff position reported that they had received support from NGOs.

When comparing experiences of team communication among health facilities, those comprising only HEWs had the lowest proportion of team meetings about service statistics in the past 12 months (n=60, 71.4%). Overall, 80.7% (n=188) of facilities with no HEWs and 89.1% (n=98) of those with at least one HEW had meetings to discuss service statistics for family planning in the past 12 months.

Disparities in Team Meetings About Service Statistics

Table 2 shows the results from the logistic regression analysis that examined the association between team communication on family planning service statistics and the representation of HEWs in staffing structures. For health facilities with HEWs as the only staff, the odds of having team meetings on family planning service statistics during the last 12 months were 0.48 times compared to those without HEWs (*P*=.02). However, there was no statistically significant difference between HEW-only facilities and facilities with at least one HEW plus other staff. Figure 1 presents the estimated predicted probabilities of having team communication among facilities with different levels of HEW integration in their staffing structure.

The association between knowledge on catchment population size and team communication experience was as follows: facilities that were aware of the size of the catchment area were more likely to report instances of team communication regarding service statistics in the last 12 months compared to those that did not have a catchment area or had no knowledge of its size (odds ratio [OR] 2.21, *P*=.03).

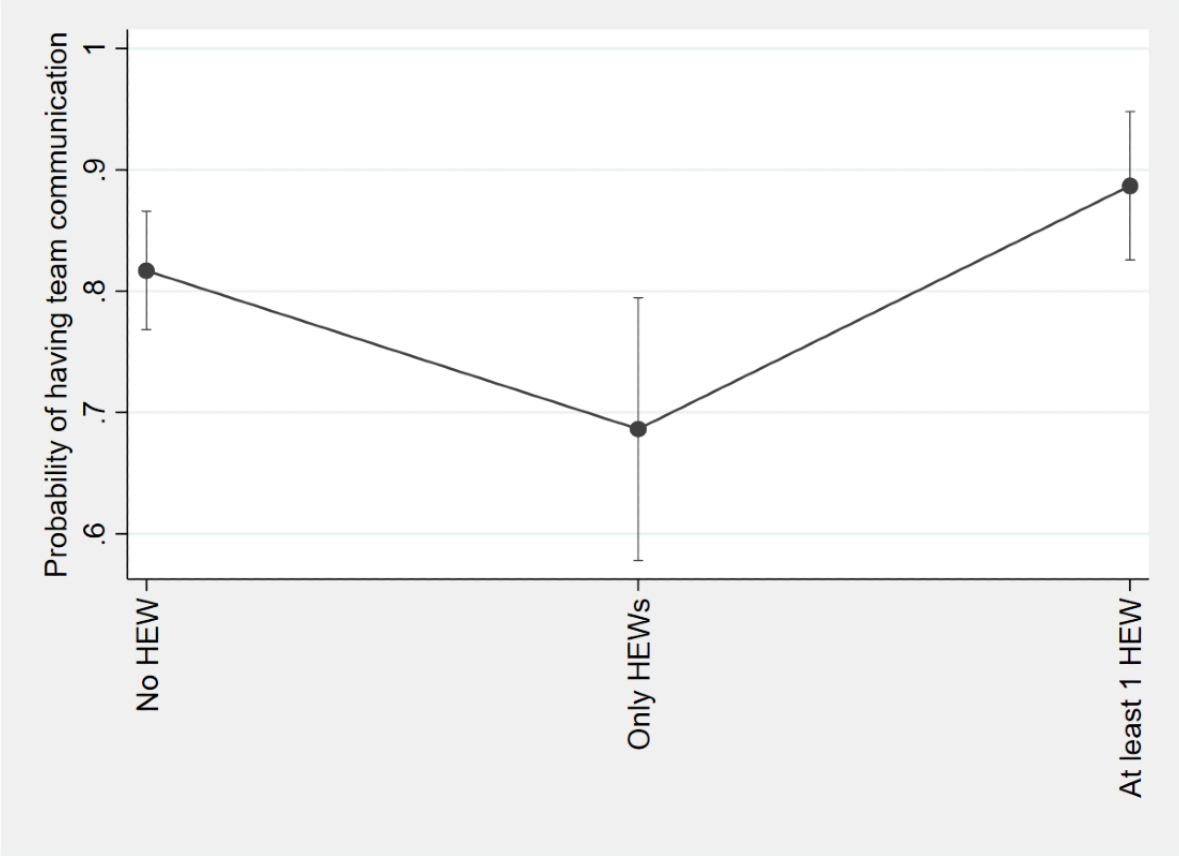
Table 2. Logistic regression model for the association between team communication on family planning service statistics and the integration of health extension workers (HEWs) into the staffing structure (source: PMA2020 Ethiopia Round 6 Service Delivery Point 2018 [19]).

Variable	Team communication in the past 12 months		
	Odds ratio (95% CI)	SE	*P* value
Integration of HEWs into staffing structure (ref[a]: no HEWs)			
At least one HEW	1.78 (0.88-3.62)	0.65	.11
Only HEWs	0.48 (0.25-0.90)	0.15	.02
Location (ref: rural)			
Urban	0.68 (0.39-1.18)	0.19	.17
Knowledge of catchment population size (ref: none or no catchment area)			
Yes	2.21 (1.06-4.64)	0.84	.03
Recent external supervisor visit (ref: >6 months ago)			
Within the past 6 months	1.02 (0.46-2.25)	0.41	.96
Received support in the last 12 months from an NGO[b] for family planning services (ref: no)			
Yes	1.47 (0.88-2.45)	0.39	.14

[a]ref: reference.

[b]NGO: nongovernmental organization.

Figure 1. Predicted probabilities of having team communication in the past 12 months among facilities with different levels of health extension worker (HEW) integration.



Disparities in Access to Electricity

Table 3 illustrates the disparities in access to electricity among health facilities with no HEWs, at least one HEW, and only HEWs in their staffing structure. When asked if they had electricity at the time of the survey, the majority of the respondents from facilities with no HEWs (190/233, 81.55%) and those with at least one HEW (88/110, 80.0%) said that they had access to electricity. On the contrary, 83.13% (69/83) of the respondents from health facilities with only HEWs responded that they had no electricity at the time of the survey. Chi-square test results indicate that there was a statistically significant difference in this regard (P<.001).

Table 3. Disparities in access to electricity for health facilities with different levels of health extension worker (HEW) integration into their staffing structure.

Variable	Access to electricity		*P* value[a]
	No, n (%)	Yes, n (%)	
Intermittent electricity access (ie, disruption lasting 2 or more hours)			.001
No HEWs	156 (66.95)	77 (33.05)	
At least one HEW	70 (64.22)	39 (35.78)	
Only HEWs	23 (28.75)	57 (71.25)	
Has electricity at this time			.001
No HEWs	43 (18.45)	190 (81.55)	
At least one HEW	22 (20.00)	88 (80.00)	
Only HEWs	69 (83.13)	14 (16.87)	

[a]Chi-square test.

However, when the respondents were asked if they had electricity at least on the day of the survey, but with a service disruption of 2 or more hours, 71.25% (n=57) of facilities with only HEWs in their staffing structure said “yes.” For facilities with no HEWs and those with at least one HEW, the percentages were much lower: 33.05% (n=77) and 35.78% (n=39), respectively. Again, the chi-square test result indicated that there was a statistically significant difference in intermittent access to electricity among health facilities with different levels of HEW integration in their staffing structure (P<.001).

Discussion

Principal Findings

This study demonstrated that health facilities in Ethiopia with only HEWs are less likely to have team meetings to discuss service statistics and more likely to experience unstable access to electricity than those without HEWs. To further explore the gaps in team communication, we focused on possible challenges among HEWs in terms of scheduling a meeting in a physical location and generating basic statistics from the paper-based field books they carry while traveling.

First of all, convening a meeting at a physical location may not be feasible for HEWs who travel a lot. Given the HEWs' role as CHWs, the majority of their working time is devoted to outreach activities and extensive travel [12]. As evidenced by previous studies, they spend approximately 75% of their working time on outreach activities and 15.5% on travel [8,10-12]. We confirmed with one health bureau staff that, in general, 80% of the HEW's working time or 4 days per week is spent on visiting households or conducting outreach activities. Consequently, HEWs may face difficulties in meeting team members to discuss service statistics, and the results from this study provide evidence on gaps between facilities with only HEWs and those without HEWs.

Another important aspect of these gaps in team communication on service statistics is the extra burden involved to generate basic statistics. As previous studies have stated, part of the reason for these gaps may come from the inefficiencies associated with the paper-based recordkeeping practice among HEWs [13,27]. When HEWs travel for their outreach activities, they usually carry a paper-based field book for recording purposes. In this field book, they record information about clients as well as services provided. After they return from the community, HEWs spend extra hours updating various information sheets for reporting, such as tally sheets, kebele (lowest administrative unit) profiling templates, or follow-up cards for various types of health conditions [8]. In addition to the inefficiencies associated with paper-based records, there is a risk for loss, duplication, or invalidity of data. These challenges, which are related to service statistics, can explain the gaps identified in this study to some extent.

Building upon previous studies, we considered mobile communication and data collection as a potential solution for these challenges [17]. In doing so, we identified disparities in access to electricity, which are potential barriers to mobile communication and data collection adoption. The disparities in access to electricity identified from this study support the need for and the feasibility of this approach. According to our results, health facilities with HEWs as the only staff had a substantially lower level of access to electricity at the time of the survey (16.87%) compared to facilities with no HEWs (81.55%) or with at least one HEW (80.0%). However, it was noted that 71.25% of facilities with only HEWs had intermittent access to electricity, implying the possibility of using it for charging mobile devices between travels.

Ideally, stable access to electricity and wireless networks would be needed for the effective use of mobile devices among HEWs. Currently, there is a pilot program involving an electronic community health information system for some health posts in the Oromia region so that HEWs can use a tablet computer for recordkeeping purposes. In addition, previous studies on HEWs in Ethiopia have already examined the potential for and the feasibility of using mobile devices for communication and data collection [13,28-31]. A qualitative study based on the Ethiopian context suggested that HEWs who participated in the intervention involving smartphones for data collection saw it as a "helping hand" [31].

Our study has limitations in terms of generalizability and the variables available from the secondary data. First, the scope of the analysis is focused on HEWs in Ethiopia although CHWs are deployed across many LMICs. The rationale for choosing Ethiopia as a target for this exploratory analysis was based on its well-structured staffing model for primary health care at the community level involving CHWs employed by the government. Future studies can explore the gaps in team communication on service statistics among CHWs in other contexts.

Secondly, the analysis is limited to the available variables from the PMA2020 data, which is secondary data collected by a global initiative. Most importantly, team communication for sharing information on service statistics can be assessed by various measures but the PMA2020 data provided only information on whether the facility had staff meetings on service statistics for family planning for the past 12 months or not. Also, additional measures that can examine facilitators and barriers for mobile communication and data collection should be examined in addition to electricity access. Based on our findings, future studies can investigate other variables that can explain the gaps in team communication on service statistics and the potential factors for assessing the feasibility of mobile communication and data collection.

As an exploratory analysis, this study attempted to provide empirical evidence on gaps in team communication about service statistics among CHWs in Ethiopia and explore the potential for mobile communication and data collection. As evidenced by current pilot programs and previous literature, utilization of mobile communication and data collection by HEWs can help bridge the gaps in team communication concerning service statistics.

Conclusions

This study aimed to examine the gaps in team communication for discussing service statistics among HEWs in Ethiopia. On the one hand, HEWs face challenges in convening for meetings at a physical location due to the extensive outreach activities and travel associated with their line of work. On the other hand, generating basic statistics from the paper-based field books that HEWs carry while traveling requires extra effort. In this context, we explored the potential for mobile communication and data collection and discussed how it can transform the way that HEWs cross-check and share service statistics data. Building upon the findings from this study, the workflow of HEWs can be improved for efficient team communication on service statistics.

Acknowledgments

SL was involved in the conceptualization of the study and drafted the manuscript. EK collected data and conducted the literature review. TBD collected information about the local context and interpreted the findings. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Dynes M, Stephenson R, Hadley C, Sibley L. Factors shaping interactions among community health workers in rural Ethiopia: rethinking workplace trust and teamwork. J Midwifery Womens Health 2014 Jan;59 Suppl 1(s1):S32-S43 [FREE Full text] [doi: 10.1111/jmwh.12135] [Medline: 24588914]
2. Franklin CM, Bernhardt JM, Lopez RP, Long-Middleton ER, Davis S. Interprofessional Teamwork and Collaboration Between Community Health Workers and Healthcare Teams: An Integrative Review. Health Serv Res Manag Epidemiol 2015;2:2333392815573312 [FREE Full text] [doi: 10.1177/2333392815573312] [Medline: 28462254]
3. Yeboah-Antwi K, Snetro-Plewman G, Waltensperger KZ, Hamer DH, Kambikambi C, MacLeod W, et al. Measuring teamwork and taskwork of community-based "teams" delivering life-saving health interventions in rural Zambia: a qualitative study. BMC Med Res Methodol 2013 Jun 27;13(1):84 [FREE Full text] [doi: 10.1186/1471-2288-13-84] [Medline: 23802766]
4. Pallangyo E, Mbekenga C, Olsson P, Eriksson L, Bergström A. Implementation of a facilitation intervention to improve postpartum care in a low-resource suburb of Dar es Salaam, Tanzania. Implement Sci 2018 Jul 28;13(1):102 [FREE Full text] [doi: 10.1186/s13012-018-0794-x] [Medline: 30055638]
5. Mishra A. 'Trust and teamwork matter': community health workers' experiences in integrated service delivery in India. Glob Public Health 2014;9(8):960-974 [FREE Full text] [doi: 10.1080/17441692.2014.934877] [Medline: 25025872]
6. WHO Guideline on Health Policy and System Support to Optimize Community Health Worker Programmes. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/275474/9789241550369-eng.pdf?ua=1 [accessed 2020-05-16]
7. Jacquelyn CA, Langer A. Health Extension Workers in Ethiopia. Maternal Health Task Force, Harvard School of Public Health, Integrare. 2014. URL: https://cdn2.sph.harvard.edu/wp-content/uploads/sites/32/2014/09/HSPH-Ethiopia4.pdf [accessed 2020-05-16]
8. Desta FA, Shifa GT, Dagoye DW, Carr C, Van Roosmalen J, Stekelenburg J, et al. Identifying gaps in the practices of rural health extension workers in Ethiopia: a task analysis study. BMC Health Serv Res 2017 Dec 20;17(1):839 [FREE Full text] [doi: 10.1186/s12913-017-2804-0] [Medline: 29262806]
9. Mangham-Jefferies L, Mathewos B, Russell J, Bekele A. How do health extension workers in Ethiopia allocate their time? Hum Resour Health 2014 Oct 14;12:61 [FREE Full text] [doi: 10.1186/1478-4491-12-61] [Medline: 25315425]
10. Medhanyie A, Spigt M, Kifle Y, Schaay N, Sanders D, Blanco R, et al. The role of health extension workers in improving utilization of maternal health services in rural areas in Ethiopia: a cross sectional study. BMC Health Serv Res 2012 Oct 08;12(1):352 [FREE Full text] [doi: 10.1186/1472-6963-12-352] [Medline: 23043288]
11. Kok MC, Ormel H, Broerse JEW, Kane S, Namakhoma I, Otiso L, et al. Optimising the benefits of community health workers' unique position between communities and the health sector: A comparative analysis of factors shaping relationships in four countries. Glob Public Health 2017 Nov 30;12(11):1404-1432. [doi: 10.1080/17441692.2016.1174722] [Medline: 27133127]
12. Tilahun H, Fekadu B, Abdisa H, Canavan M, Linnander E, Bradley EH, et al. Ethiopia's health extension workers use of work time on duty: time and motion study. Health Policy Plan 2017 May 01;32(3):320-328. [doi: 10.1093/heapol/czw129] [Medline: 27658649]
13. Medhanyie AA, Moser A, Spigt M, Yebyo H, Little A, Dinant G, et al. Mobile health data collection at primary health care in Ethiopia: a feasible challenge. J Clin Epidemiol 2015 Jan;68(1):80-86. [doi: 10.1016/j.jclinepi.2014.09.006] [Medline: 25441699]
14. Kare CT. Community Helath Information System for Family-Centered Health Care: Scale-Up in Southern Nations, Nationalities and People's Region. Federal Democratic Republic of Ethiopia, Ministry of Health, Quarterly Health Bulletin. 2013. URL: https://www.measureevaluation.org/resources/publications/ja-13-161 [accessed 2020-05-16]
15. Hirvonen K, Berhane G, Assefa TW. Assessing Community Health Information Systems: Evidence from Child Health Records in Food Insecure Areas of the Ethiopian Highlands. Matern Child Health J 2020 Aug 28;24(8):1028-1037 [FREE Full text] [doi: 10.1007/s10995-020-02943-1] [Medline: 32347438]
16. Assefa Y, Gelaw YA, Hill PS, Taye BW, Van Damme W. Community health extension program of Ethiopia, 2003-2018: successes and challenges toward universal coverage for primary healthcare services. Global Health 2019 Mar 26;15(1):24 [FREE Full text] [doi: 10.1186/s12992-019-0470-1] [Medline: 30914055]

17. Martin G, Khajuria A, Arora S, King D, Ashrafian H, Darzi A. The impact of mobile technology on teamwork and communication in hospitals: a systematic review. J Am Med Inform Assoc 2019 Apr 01;26(4):339-355 [FREE Full text] [doi: 10.1093/jamia/ocy175] [Medline: 30689893]
18. White A, Thomas DSK, Ezeanochie N, Bull S. Health Worker mHealth Utilization: A Systematic Review. Comput Inform Nurs 2016 Mar 7. [doi: 10.1097/CIN.0000000000000231] [Medline: 26955009]
19. PMA2020. PMA Data. 2020. URL: https://www.pmadata.org/data [accessed 2020-05-16]
20. Zimmerman L, Olson H, PMA2020 Principal Investigators Group, Tsui A, Radloff S. PMA2020: Rapid Turn-Around Survey Data to Monitor Family Planning Service and Practice in Ten Countries. Stud Fam Plann 2017 Sep;48(3):293-303 [FREE Full text] [doi: 10.1111/sifp.12031] [Medline: 28885679]
21. Negussie A, Girma G. Is the role of Health Extension Workers in the delivery of maternal and child health care services a significant attribute? The case of Dale district, southern Ethiopia. BMC Health Serv Res 2017 Oct 11;17(1):641 [FREE Full text] [doi: 10.1186/s12913-017-2590-8] [Medline: 28893267]
22. Shah NM, Wang W, Bishai DM. Comparing private sector family planning services to government and NGO services in Ethiopia and Pakistan: how do social franchises compare across quality, equity and cost? Health Policy Plan 2011 Jul 05;26 Suppl 1(Suppl. 1):i63-i71 [FREE Full text] [doi: 10.1093/heapol/czr027] [Medline: 21729919]
23. Olson DJ, Piller A. Ethiopia: an emerging family planning success story. Stud Fam Plann 2013 Dec 09;44(4):445-459. [doi: 10.1111/j.1728-4465.2013.00369.x] [Medline: 24323662]
24. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
25. Kok MC, Kea AZ, Datiko DG, Broerse JEW, Dieleman M, Taegtmeyer M, et al. A qualitative assessment of health extension workers' relationships with the community and health sector in Ethiopia: opportunities for enhancing maternal health performance. Hum Resour Health 2015 Oct 30;13:80 [FREE Full text] [doi: 10.1186/s12960-015-0077-4] [Medline: 26423049]
26. Sedlander E, Bingenheimer JB, Edberg MC, Rimal RN, Shaikh H, Munar W. Understanding modern contraception uptake in one Ethiopian community: a case study. Reprod Health 2018 Jul 20;15(1):111 [FREE Full text] [doi: 10.1186/s12978-018-0550-3] [Medline: 29925395]
27. Feroz A, Jabeen R, Saleem S. Using mobile phones to improve community health workers performance in low-and-middle-income countries. BMC Public Health 2020 Jan 13;20(1):49 [FREE Full text] [doi: 10.1186/s12889-020-8173-3] [Medline: 31931773]
28. Mengesha W, Steege R, Kea AZ, Theobald S, Datiko DG. Can mHealth improve timeliness and quality of health data collected and used by health extension workers in rural Southern Ethiopia? J Public Health (Oxf) 2018 Dec 01;40(suppl_2):ii74-ii86 [FREE Full text] [doi: 10.1093/pubmed/fdy200] [Medline: 30551131]
29. Damtew Z. Accelerating health service and data capturing through community health workers in rural Ethiopia: A pre-requisite to progress. 2012 Presented at: KMIS 2012- Proceedings of the International Conference on Knowledge Management and Information Sharing; 4-7 October, 2012; Barcelona, Spain p. 168-177 URL: https://pdfs.semanticscholar.org/5970/d801113501efb3adcfb45a9e8745c40c31e6.pdf [doi: 10.5220/0004135101680177]
30. Little A, Medhanyie A, Yebyo H, Spigt M, Dinant G, Blanco R. Meeting community health worker needs for maternal health care service delivery using appropriate mobile technologies in Ethiopia. PLoS One 2013 Oct 29;8(10):e77563 [FREE Full text] [doi: 10.1371/journal.pone.0077563] [Medline: 24204872]
31. Steege R, Waldman L, Datiko DG, Kea AZ, Taegtmeyer M, Theobald S. 'The phone is my boss and my helper' - A gender analysis of an mHealth intervention with Health Extension Workers in Southern Ethiopia. J Public Health (Oxf) 2018 Dec 01;40(suppl_2):ii16-ii31 [FREE Full text] [doi: 10.1093/pubmed/fdy199] [Medline: 30551130]

Abbreviations

CHW: community health worker
HEW: health extension worker
LMIC: low- and middle-income country
NGO: nongovernmental organization
OR: odds ratio
PMA2020: Performance Monitoring and Accountability 2020
WHO: World Health Organization

Edited by G Eysenbach; submitted 31.05.20; peer-reviewed by M Fiordelli, J VanDerLinde; comments to author 19.06.20; revised version received 24.07.20; accepted 10.08.20; published 08.09.20.

Please cite as:
Lee S, Kim E, Desta TB
Gaps in Team Communication About Service Statistics Among Health Extension Workers in Ethiopia: Secondary Data Analysis
JMIR Mhealth Uhealth 2020;8(9):e20848
URL: http://mhealth.jmir.org/2020/9/e20848/
doi:10.2196/20848
PMID:32897231

Original Paper

Cost-Effective Smartphone-Based Articulable Endoscope Systems for Developing Countries: Instrument Validation Study

Youngjin Moon[1,2*], PhD; Jeongmin Oh[1*], BSc; Jaeho Hyun[1*], MSc; Youngkyu Kim[1], MSc; Jaesoon Choi[1,3], PhD; Jeongman Namgoong[4], MD, PhD; Jun Ki Kim[1,2], PhD

[1]Biomedical Engineering Research Center, Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea

[2]Department of Convergence Medicine, College of Medicine, University of Ulsan, Seoul, Republic of Korea

[3]Department of Biomedical Engineering, College of Medicine, University of Ulsan, Seoul, Republic of Korea

[4]Department of Surgery, Asan Medical Center, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jun Ki Kim, PhD
Biomedical Engineering Research Center
Asan Institute for Life Sciences
Asan Medical Center
88, Olympic-ro 43-gil, Songpa-Gu
Seoul,
Republic of Korea
Phone: 82 23 010 8619
Fax: 82 23 010 4182
Email: kim@amc.seoul.kr

Abstract

Background: Endoscopes are widely used for visualizing the respiratory tract, urinary tract, uterus, and gastrointestinal tracts. Despite high demand, people in underdeveloped and developing countries cannot obtain proper access to endoscopy. Moreover, commercially available endoscopes are mostly nonarticulable as well as not actively controlled, limiting their use. Articulating endoscopes are required for some diagnosis procedures, due to their ability to image wide areas of internal organs. Furthermore, actively controlled articulating endoscopes are less likely to harm the lumen than rigid endoscopes because they can avoid contact with endothelial tissues.

Objective: The study aimed to demonstrate the feasibility and acceptability of smartphone-based wide-field articulable endoscope system for minimally invasive clinical applications in developing and less developed countries.

Methods: A thin articulable endoscope system that can be attached to and actively controlled by a smartphone was designed and constructed. The system consists of a flexible endoscopic probe with a continuum mechanism, 4 motor modules for articulation, a microprocessor for controlling the motor with a smartphone, and a homebuilt app for streaming, capturing, adjusting images and video, and controlling the motor module with a joystick-like user interface. The smartphone and motor module are connected via an integrated C-type On-The-Go (OTG) USB hub.

Results: We tested the device in several human-organ phantoms to evaluate the usability and utility of the smartphone-based articulating endoscope system. The resolution (960 × 720 pixels) of the device was found to be acceptable for medical diagnosis. The maximum bending angle of 110° was designed. The distance from the base of the articulating module to the tip of the endoscope was 45 mm. The angle of the virtual arc was 40.0°, for a curvature of 0.013. The finest articulation resolution was 8.9°. The articulating module succeeded in imaging all 8 octants of a spherical target, as well as all 4 quadrants of the indices marked in human phantoms.

Conclusions: The portable wide-field endoscope was successfully controlled using a smartphone, yielding clear images with a resolution of 960 × 720 pixels at realistic focal distances. Actively and precisely controlled articulating movements have resulted in minimally invasive monitoring in the narrow space of internal organs providing a wide-area view. We found our smartphone-based active articulated endoscope to be suitable for point-of-care applications in developing and less developed countries.

(JMIR Mhealth Uhealth 2020;8(9):e17057) doi:10.2196/17057

KEYWORDS

smartphone-based endoscope; mobile health; low-resource settings; continuum body; articulable endoscope; low-cost medical device; point of care diagnostics

Introduction

Over the past few decades, endoscopy has become a fundamental clinical approach for noninvasive screening and diagnosis of various diseases. Now, endoscopy designers have expanded the range of functions to include longitudinal observation in noninvasive methods, simple medical procedures, biopsy, local therapy, or minimally invasive robotic surgery [1,2]. Despite their essential utilities for screening and diagnosis, clinical endoscopes rarely leave the hospital; thus, they are used mainly for clinical applications such as biopsy and staging for diseases.

In resource-poor undeveloped and developing countries, it is difficult to obtain proper endoscopy service. For example, the majority of hospitals in Nigeria have no facilities for endoscopy [3]. Moreover, less developed countries often have serious power supply problems. The medical power supply does not satisfy the needs of patients; consequently, proper medical care relies on local generators. The poor supply of power has deleterious effects on medical endoscopy, because endoscopic procedures are abruptly stopped due to power interruptions [3]. In addition to power deficiencies, several other problems in developing countries disrupt endoscopy service to serious levels. The limited number of hospitals that are able to provide endoscopy services, poor maintenance, and broken endoscopes, as well as the long distances between rural areas and large hospitals, combine to prevent people in need from receiving proper endoscopy service.

Despite the increasing demands for overcoming current problems, there have been few attempts to integrate medical devices into smartphones having ample potential for solving those limitations. In addition to having the performance of a personal computer, smartphones are portable and can be equipped with many useful peripherals, including selfie sticks, USB devices, sensors, and supplementary microprocessors. However, most of the apps of smartphone devices are based on the sensors built into smartphones, such as the cameras, accelerometer, gyroscope, magnetometer, and barometer. Those functions are commonly used to determine the user's position, movement, and posture. Some of the smartphone attachments are developed for biological analysis with simple optical devices which enable to perform enzyme-linked immunosorbent assay analysis [4] and immunoassays or to collect and analyze recombinant bovine somatotropin microsphere assay data of food allergens [5].

There are various smartphone-based endoscopes on the market. These endoscopes are small, long webcams that can be connected to smartphones via USBs that enable endoscopic webcams to be connected to desktops, laptops, and even smartphones [6]. The price range of these peripheral devices varies from US $10 to $100 [7]. These peripheral devices are employed for industrial applications such as pipe and car inspection. They are mostly flexible to various degrees, including flexible rubber covering inner electronics and a long spine-like exoskeleton that can be fixed in a position when shaped by external forces; however, the flexible probe cannot be actively controlled. These passively bendable endoscopes have a long, stiff top-head and large diameter due to their exoskeletal rubber jacket, making it difficult to access narrow spaces. Endoscopes with a spine-like exoskeleton are not only extremely thick to be employed for medical examination but also extremely rigid, thus presenting a risk of harming the soft inner surface of the organs [8]. Therefore, neither of these endoscopes are suitable for medical inspection. Moreover, none of the commercially available smartphone endoscopes are actively articulable; thus, they cannot actively scan tubular organ canals. A few rigid endoscopes are designed such that they can inspect the insides of the human body; however, they are rigid and short, limiting them to the shallow parts of the body (eg, the nostrils and entrance to the ear canals).

The articulating endoscope is needed in medical diagnosis for several reasons. The first is patient safety and comport. Patient comfort following a minimally invasive procedure differs significantly with flexible versus rigid endoscope use. In addition, actively articulated flexible endoscopes can provide better safety and comfort than passively articulated ones. Gmeiner et al [9] reported that flexible endoscopy resulted in more patients being comfortable after treatment, while rigid endoscopy always required general anesthesia or sedation during the procedure. For example, not all patients tolerate rigid laryngoscopy, particularly those with a sensitive gag reflex or limited jaw or neck mobility [10]. Sensitive patients or patients with limited movements undergoing rigid laryngoscopy are more likely to feel uncomfortable during and after the procedure. Additionally, patients reported less pain during cystoscopy performed with flexible cystoscopes compared with rigid cystoscopes [11,12]. The second reason why articulating endoscopes are needed in medical diagnosis is that they can provide wide-area monitoring. Since the intestinal canal or cavities in the human body have mostly abrupt bending and curves, the rigidity of the endoscope significantly influences the range of view that the medical professional can observe in the organ. Unlike rigid endoscopes, articulating endoscopes achieved an imaging success rate of 90% and covered almost all ranges and directions as long as they had proper bending angles [13,14].

In this paper, we present a design to satisfy the need for an actively controlled articulating endoscope with smartphone connectivity and mobile medical device applications, in the absence of commercially available equivalent. Our device connects to a smartphone via USB connection, and a homebuilt app allows the user to successfully adjust images, control the endoscope's articulable robotic body, and capture still images and movies for later analysis. With a supplemental microprocessor, the articulable robotic continuum body achieves motion steps of 8.9°, and the camera offers a resolution of 960 × 720 pixels at 30 fps and a depth of field of 10 to 80 mm. For

additional portability, the device may be powered entirely by the smartphone's USB port, or supplementary batteries may be used. The designed device is easily cleanable and self-illuminated and attains a wide field of view and suitable resolution for clinical imaging in phantoms of the human larynx and bladder. We hope that it will provide a strong development base to satisfy clinical needs in underdeveloped countries.

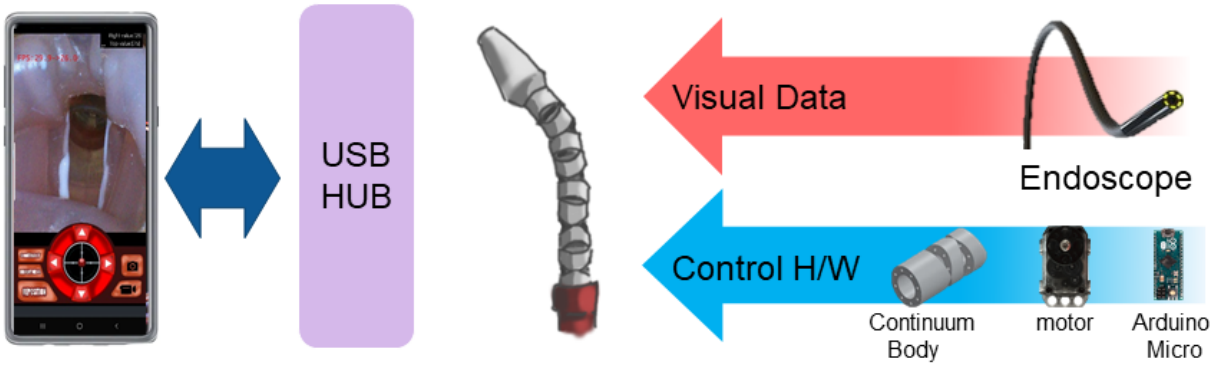
Methods

USB-Integrated Endoscope for Smartphone Appendage

The articulating and portable endoscope device was designed for the smartphone by integrating 8 parts: smartphone, four servomotors, microprocessor, articulating continuum robot tip module, wire-based system for tip articulation, USB connection circuit, case for smartphone mounting, and commercially available industrial endoscope module. The endoscope module (3.9 mm-diameter medical endoscope module with a USB interface, 100° large field of view model, Shanghai Chenyu Smart Technology Co, Ltd) supports USB 3.0 connection, features light-emitting diode illumination, and is compatible with Intel USB chips. With a diameter of 3.9 mm and a head length of 10.6 mm, the USB endoscope module is smaller in diameter than commercially available industrial endoscopes, and its thickness and length are suitable for navigating and imaging the esophagus, larynx [15], and bladder [16], which are frequently examined. The complementary metal-oxide semiconductor (CMOS) sensor of the endoscope module uses the ov9734 chip from Omnivision, which supports a resolution of 960 × 720 pixels, 30 fps video streaming, LED light source, and programmatic adjustment of the brightness, contrast, white balance, and exposure. The depth of field is 10 to 80 mm, which allows a large range of imaging in internal organs.

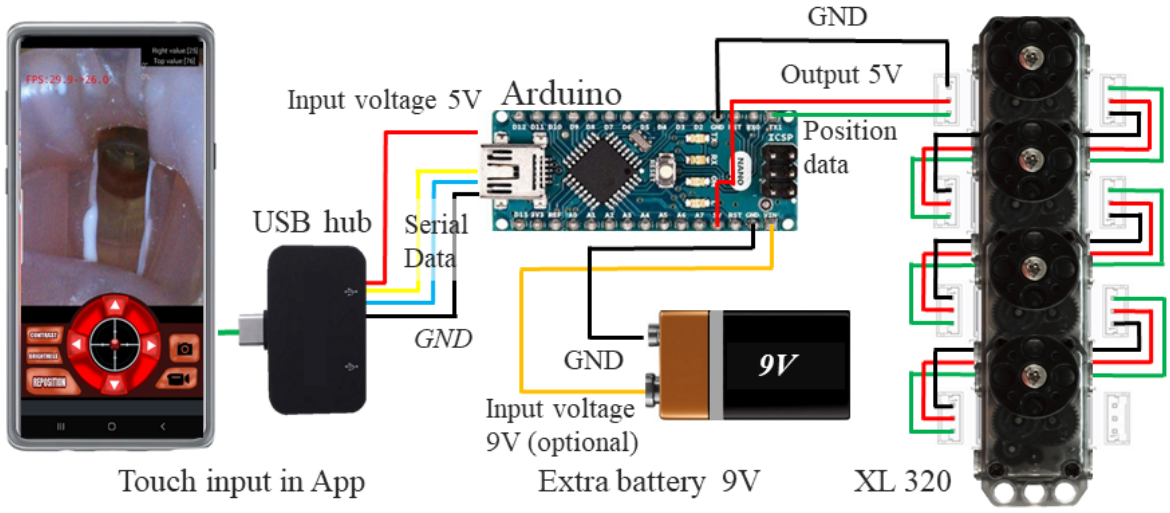
A general schematic of the endoscope is shown in Figure 1. The app functions as a software platform for the camera and for control of the snake-like continuum body. The devices are connected by an on-the-go (OTG) USB connection. USB devices intrinsically have a master/slave relationship. Only the master device can control a slave device. Smartphones are typically considered to be slave USB devices. Therefore, if connected normally, a smartphone cannot control the attached cameras and send a signal to a microprocessor. However, with the OTG USB hub connection, the roles of the master and slave can easily be changed, allowing the smartphone to serve as the host to peripheral devices [17]. A microprocessor and camera unit serve as the slave units, while the smartphone acts as a host.

Figure 1. Overall schematic of the data flow of the articulating mobile endoscope. The smartphone is attached to the articulating module, which is a combination of the endoscope module and the continuum body.



Through the OTG USB hub connection, the app, endoscope CMOS module, and microprocessor can interact with each other. Figure 2 presents a simplified circuit diagram of the device. With the OTG USB hub connection secured, the app obtains permission to access the camera and microprocessor. We used the Arduino Nano as a microprocessor, which is an open-source electronics platform, so that it can be easily used for development and circuit building [18]. The Arduino, motor, and smartphone all have the same 5V operating voltage. Extra batteries can be connected to the device if necessary, but primary electricity is supplied by the smartphone's lithium-ion battery. The OTG USB hub connection also functions as an electricity-providing hub.

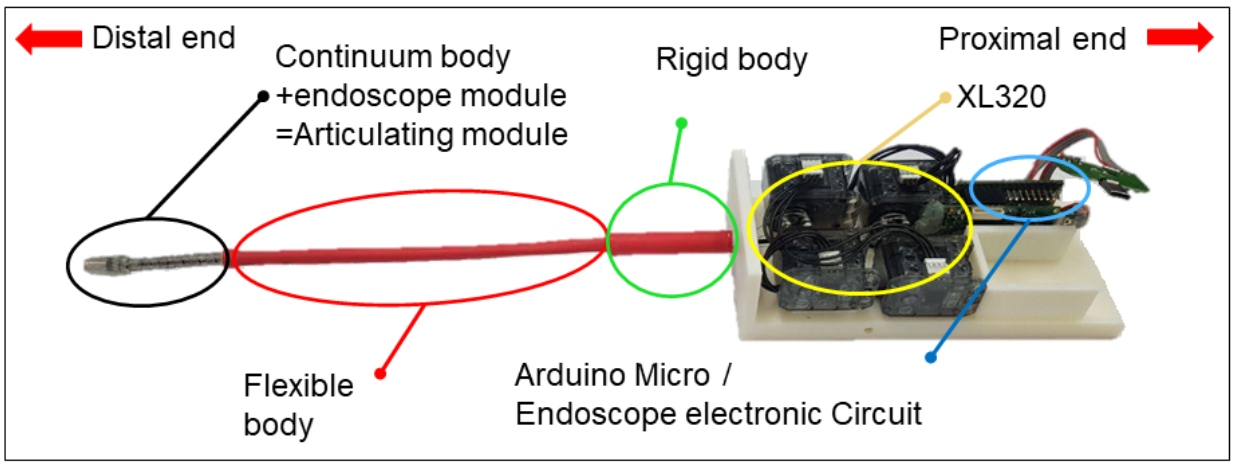
Figure 2. Electric control diagram of the articulation control system. The touch input from a smartphone app is processed into motor control signals using an Arduino processor. The extra battery is optional.



The microprocessor is connected to 4 motors, which are connected to wire-controlling pulleys. The motor (Dynamixel XL-320, Robotis) is a robot-exclusive smart actuator with a fully integrated direct current motor, reduction gearhead, controller, driver, and network in one direct current servo module [19]. Each motor has its own motor driver, allowing intuitive control by the microprocessor software.

Figure 3 depicts the assembled device prototype, complete with the 3-dimensional (3D)–printed case. The motor, microprocessor, and OTG USB hub are assembled within a proximal case. From the case, the rigid segment connected to the flexible segment protrudes. At the most distal part, the continuum robot is assembled around the endoscope to form the articulating module. Four motors are affixed to the bottom of the base of the case.

Figure 3. Cut-away assembly of the device, showing the actual structure. The long flexible distal end is composed of the continuum, flexible, and rigid bodies.



Continuum Body

The continuum body is designed to have a snake-like motion. This structure comprises several small segments connected via wires. An individual segment in the continuum body is portrayed in Figure 4A. The ring-like individual segment with 8 wire holes has internal space that can be used for essential internal components (ie, cameras); further, this space helps reduce the weight of the segment and allows for easy maintenance. The two individual segments shown in Figure 4A comprise rolling joints, and their protruding surfaces are in contact. Figure 4B shows the simulated articulation of 3 individual segments (#1,2,3) connected by control wires, using which a combination of multiple segments can move with various degrees of freedom (DOFs) [20]. Such a multijoint tendon-driven mechanism is common in medical devices because of its efficiency, robustness, and reliability [21]. The combination of the segments is articulated by applying forces to pull the control wires, which torque to the final segments about the slopes on the connecting edges [22]. The rotation axis formed by the slopes between segments #1 and #2 is perpendicular to that formed by segments #2 and #3. Therefore, a combination of 3 segments can make the arbitrary orientation of the distal segment in 2 DOFs, as shown in Figure 4B.

Figure 4. Structure of the continuum body: (a) 3D model of an individual unit of the continuum mechanism, (b) movement combination of three individual segments, resulting in 2-degree of freedom rotation, (c) bending motion of the 3D-printed continuum body prototype, and (d) prototype of the flexible and rigid bodies.

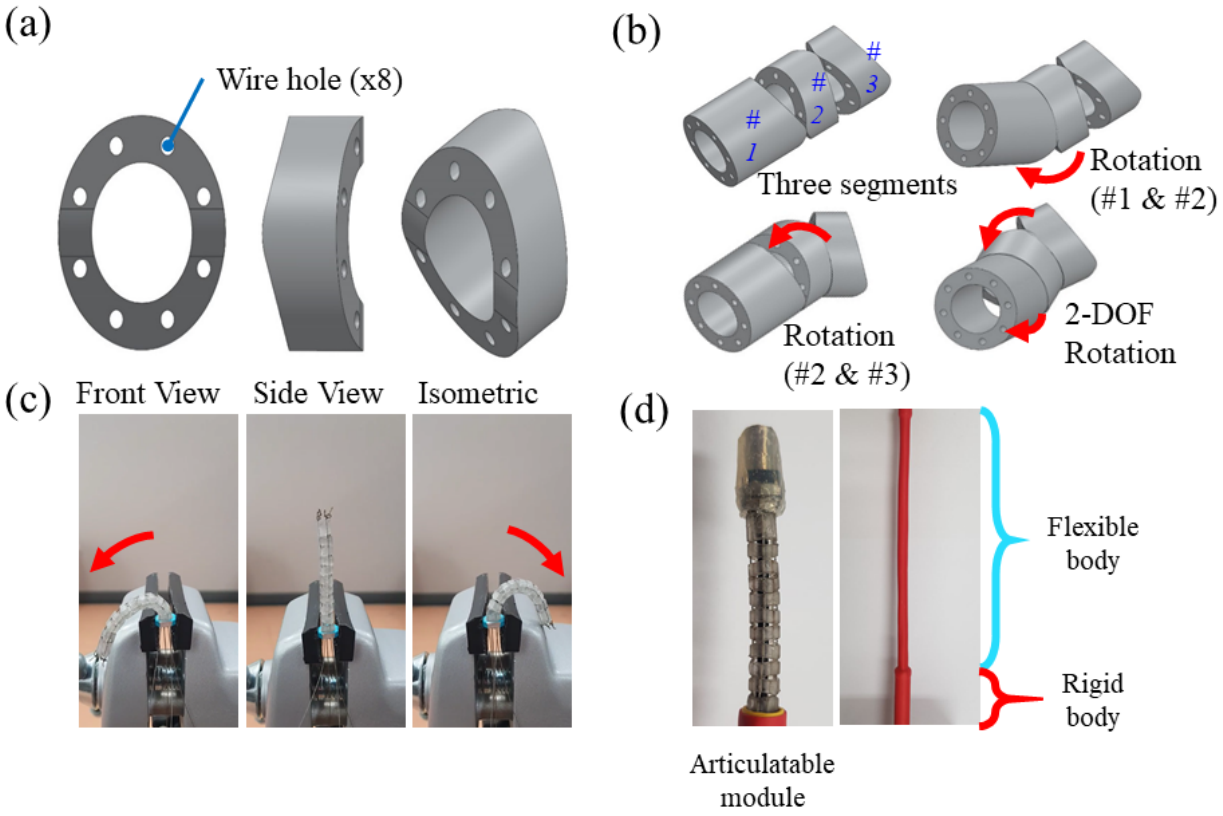
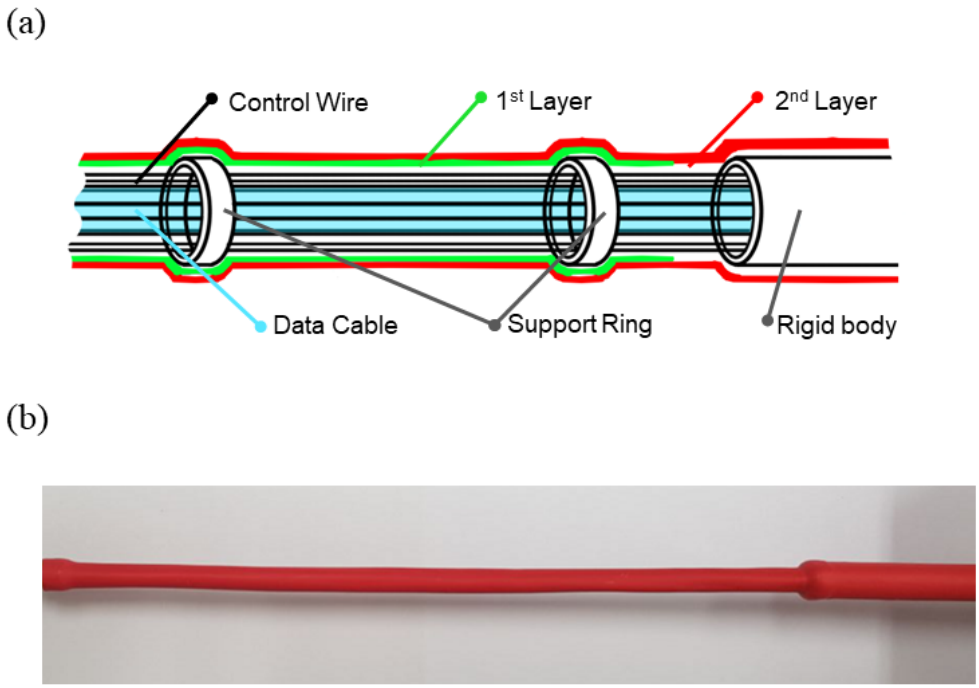


Figure 4C shows the bending of a 3D-printed continuum body prototype. The bending angle is sufficient to make the tip of the continuum body face the proximal direction. When a user touches a controlling joystick-like user interface (UI) in a smartphone app, the microprocessor controls 4 motors to tighten and loosen the pair of wires. As shown in Figure 4B, this gives the continuum mechanism 2 DOFs, which provides full freedom in movement within a 2D plane [23]. Therefore, with adjustment of the tension, the continuum mechanism can navigate through complex spatial obstacles. The model of each continuum body backbone was designed using SolidWorks 2018 (Dassault Systèmes SolidWorks Corporation) and then printed by a 3D printer.

Figure 4D shows a completed articulating module prototype, assembled with flexible and rigid bodies. The articulating module is combined with the flexible and rigid bodies to adjust the distance from a target to the controller case. The flexible body is approximately 15 cm long, allowing endoscope to smoothly slither into an organ. The rigidity of the flexible bodies is selected to suit its role. It is flexible enough to pass through canals, similar to a traditional flexible endoscope, and rigid enough for wires not to bend this part during articulation. The rigid body forms the base of the articulating wire system and is approximately 5.5 cm long.

The flexible segment and rigid segment were jacketed with a waterproof material, polyolefin, as shown in Figure 5. The jacketing process is essential for an endoscope, to protect the inner electronics from the humidity of the organ lumen and protect the patient from mechanical debris. As shown in Figure 4, the assembled segments have lateral gaps between segments. These gaps may allow mucus or substances found on the interior of an organ to enter the central hole of the articulating module, leading to circuit failure as well as injury to the vulnerable endothelium from clipping when the lateral gap diminishes.

Figure 5. Flexible segment of the endoscope module: (a) structure of the flexible segment, designed to flex to the curvature of the organ and also be stiff enough to relay the tension of the controlling wires to the articulating module and (b) fabricated flexible segment.



To prevent the flexible segment from being tensioned by the continuum segment, we applied double jacketing and added support rings for the flexible segment. Figure 5A depicts the internal structures of the flexible and rigid segments. The data cable in the center is surrounded by wires and relays tension to the continuum segment. The control wires pass through holes in the support rings. The support rings perform two major functions: preventing 8 wires from tangling in the middle of the flexible segment and allowing the flexible segment to relay sufficient force to the continuum segment, making the endoscope articulable and more responsive to the control.

While the first jacketing layer extends to the support rings, the second layer thoroughly covers the entire outer surface of the flexible segment. To ensure transparency for the articulating module, the continuum segment is jacketed in transparent latex. With double jacketing, the sheath became thick enough for the flexible segment to resist the torsional forces of the continuum segment, while remaining thin enough to fit into the sharp curvature of internal body parts. When jacketing was complete, these lateral gaps were covered by the sheath. We used a contractile tube for the sheathing and verified its feasibility in a phantom application. For clinical applications, the sheathing can be replaced with materials that are more biocompatible.

Controller Case

The controller case comprises a lid and a base, including motors and electronic boards, as shown in Figure 6. Four wires controlling the articulating module are bound to the pulleys with projecting screws installed on the rotary shafts of the respective motors. The base part is designed so that a pair of motors are mounted vertically at different heights to ensure that they are connected to each pulley without tangling from the connection hole. The remaining space of the controller case is reserved for the battery, control board, board for signal connection, etc. The lid part is designed so that a smartphone can be placed on the part, and the rear side of the case—opposite to the articulating module—is formed in an open structure so that the signal line from the controller case can be easily connected to the smartphone inlet on the upper part. The model of the case was designed using SolidWorks 2018, and the prototype was constructed via 3D printing. The dimensions of the control case are 150 mm (length), 76 mm (width), and 65 mm (height). The fabricated devices can be equipped with a 5V battery (9V extra) and a 140 mm (length) by 70 mm (width) size smartphone. The price for hardware construction are presented in Table 1.

Figure 6. Controller box: (a) configuration of the components in the controller box, (b) 3D printed case of the controller box (base of case immobilizes the motors, and top lid serves as mount for the smartphone), and (c) overall outer blueprint of assembled case and magnified image of the articulating module.

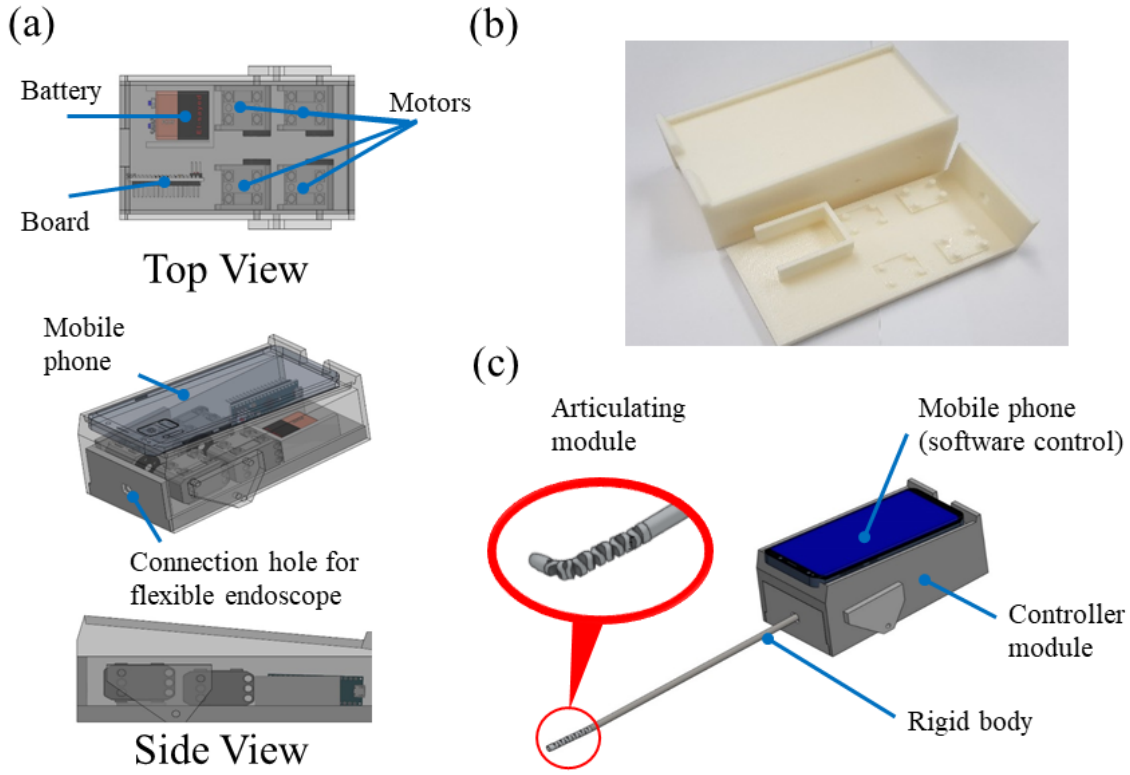


Table 1. Hardware prices of the articulable endoscope device.

Hardware	Price US$
XL320	22
3.9-mm camera module	170
3D printing of case	16
3D printing of continuum	212
Controller	19
Wires, etc.	34
Total price	473

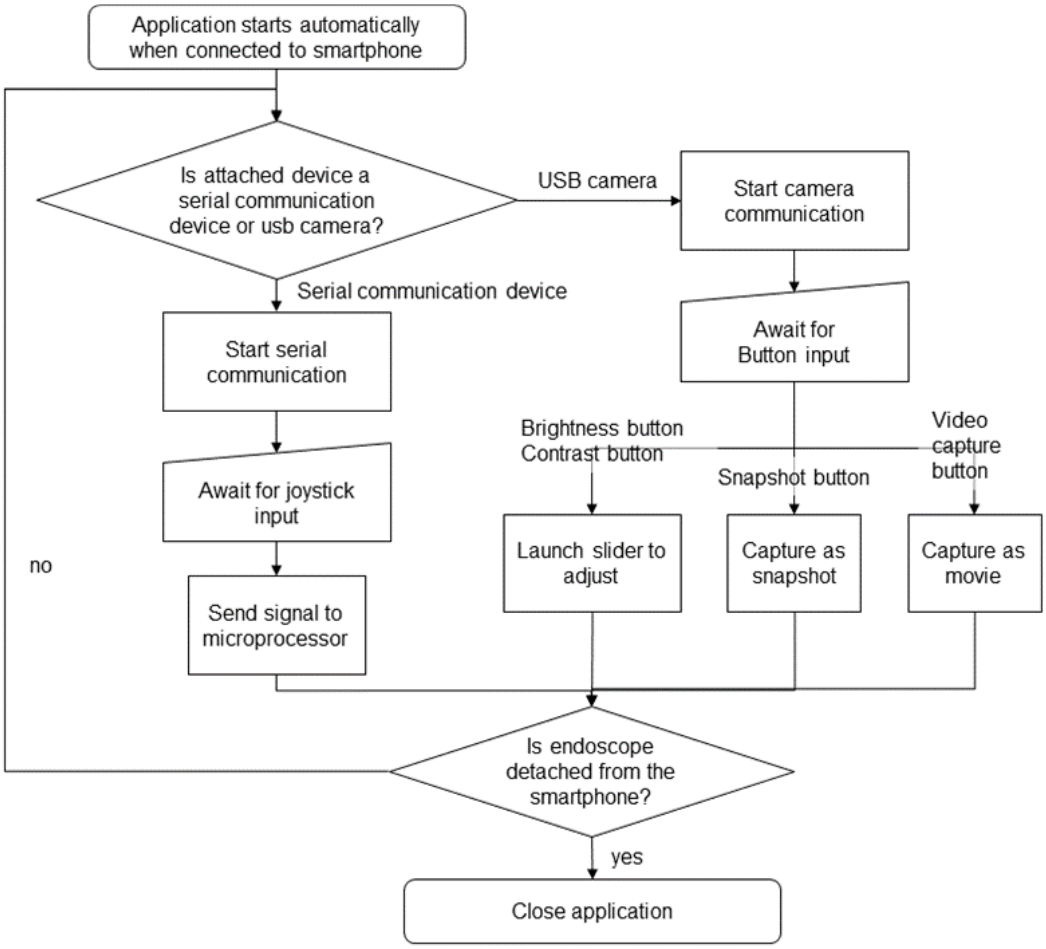
Results

User Interface and Algorithm for Smartphone App

We developed an app written in the languages Java and C to provide a straightforward and intuitive interface. The Java app included the UVC camera library developed by saki4510t [24] to enable the user to use the USB camera, as well as the USB serial library [25] for serial communication between smartphone and microprocessor. The virtual joystick library developed by Damien Brun [26] and directional buttons are employed to enable intuitive motor control.

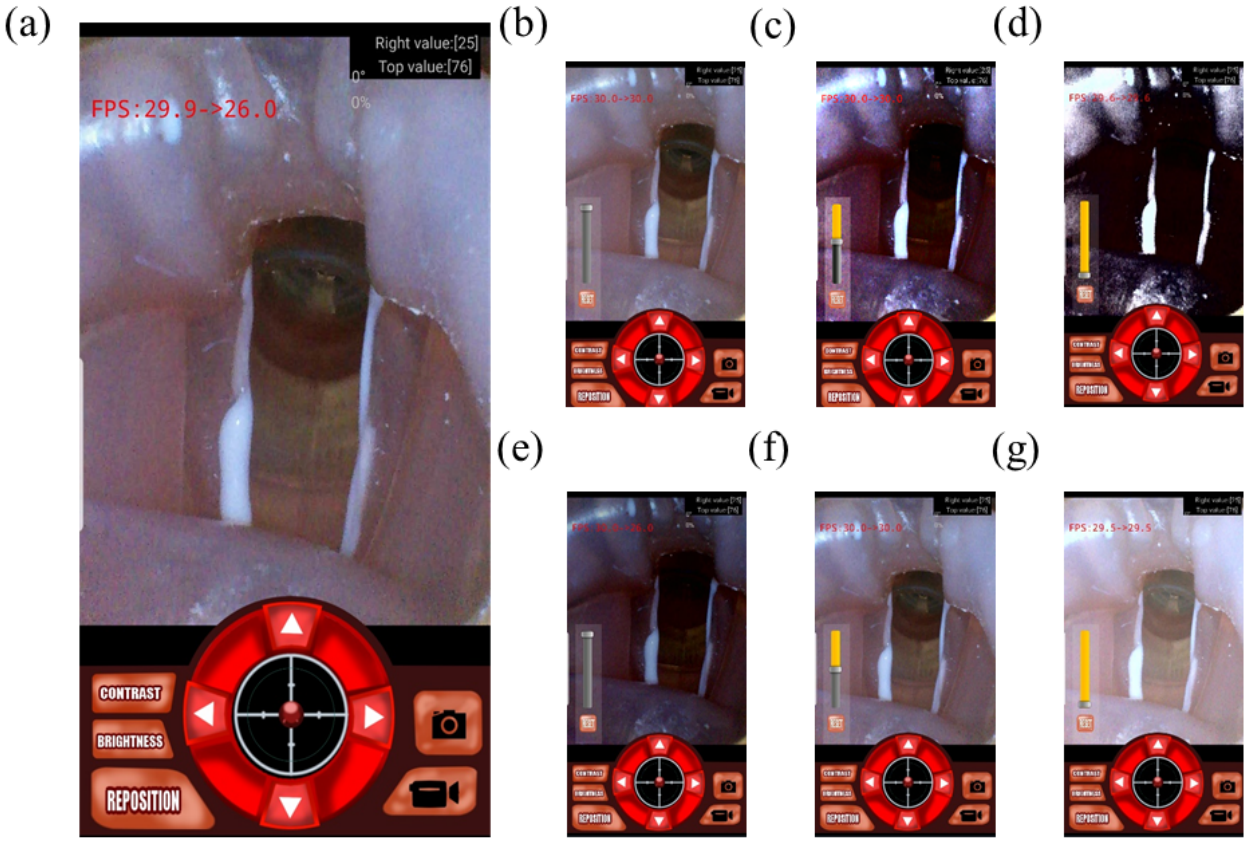
Figure 7 describes the general workflow algorithm of the app. The app automatically launches when the device is attached to the smartphone. When the OTG USB hub connection of the two device components (endoscope module and microprocessor) is secured, the app automatically distinguishes the two components of the device; one is the serial communication device (ie, the microprocessor) and the other is the CMOS chip of the endoscope camera. These two simultaneous threads allow the app to display image data from the camera while sending signals to the microprocessor. When the communication thread is launched, the app opens the intuitive UI, where the user can control the camera and motor simultaneously.

Figure 7. Overall workflow algorithm of the app.



The UI of the app is presented in Figure 8A. The app functions as a viewer, controller for the camera, motor controller, and recording device. A large center image viewer relays visual signals received by the CMOS chip in real time. The virtual joystick UI allows the user to intuitively control the motors by holding it in position, similar to the joystick of a game controller. If the user touches and drags the joystick-like controller, the app sends calculated intensity signals to the microprocessor. The microprocessor winds motors to induce movements corresponding to the virtual joystick. When the user releases the joystick-like controller, the red-dot pointer annotating the figure position returns to the origin point, and the motors stop. The 4 directional buttons are arranged around the joystick-like controller. The buttons function as a secondary motor controller, allowing the user fine control over the articulation of the endoscope. The reposition button is located on the lower left of the UI and makes the articulating module return to its origin point.

Figure 8. Software-controlling user interface (UI) of the homebuilt smartphone app: (a) main UI of the app (joystick and motor control buttons are in the center, contrast and brightness adjustment buttons are on the left, and movie and still-image capture buttons are on the right), (b-d) slider UI for contrast adjustment and corresponding processed images, and (e-g) slider UI for brightness adjustment and corresponding processed images.



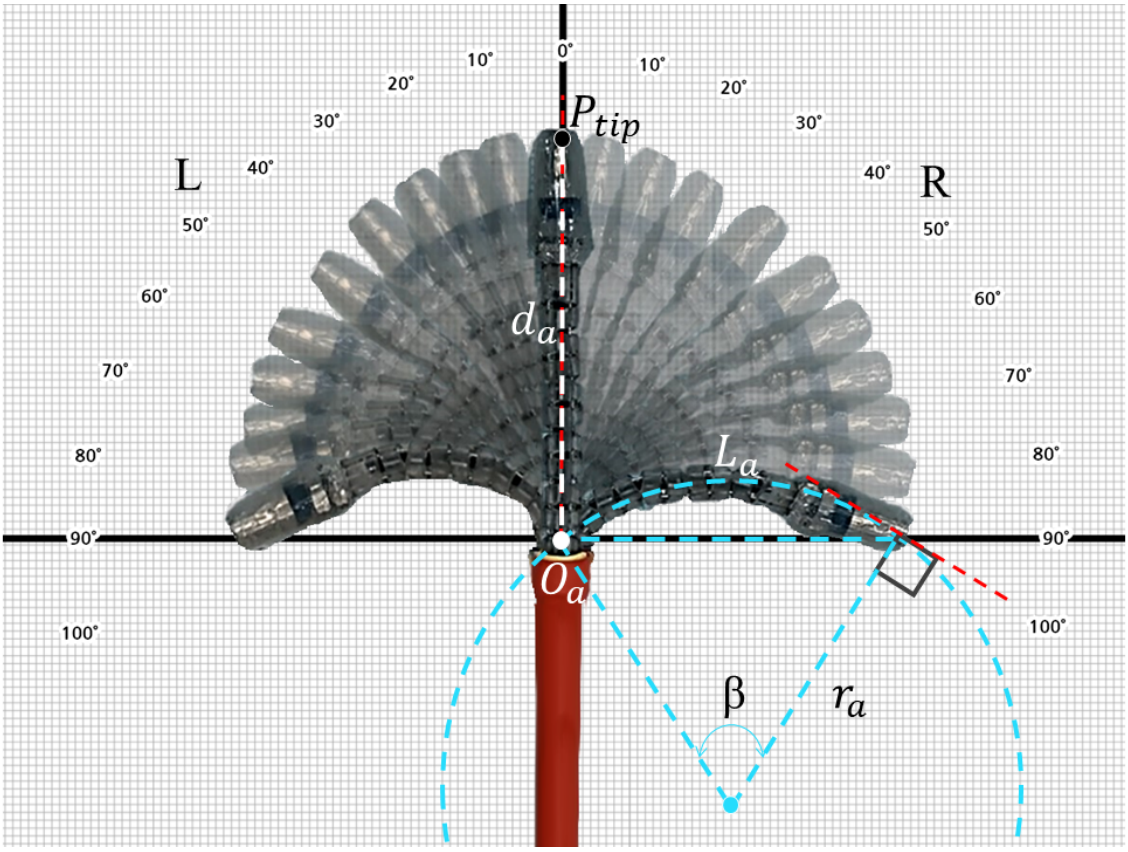
The app allows the user to capture the current view in two formats: as a still image and as a movie. The still capture button and video capture button at the bottom right of the screen perform this task. Image and videos are stored automatically in separate folders.

We arranged the button and slider UI to provide control over brightness and contrast to the user. Figures 8B-D depict the contrast adjustment slider and corresponding changes in images. If the user touches the contrast button on the bottom left of the screen, the semitransparent slider bar with the slider bar reset button appears on the screen. The user can easily adjust the contrast by touching the slider UI. The reset button resets the slider parameter to the default setting. Brightness adjustment is performed similarly. Figures 8E-G depict the brightness adjustment slider and corresponding image transformation.

Articulation Test

A bending characteristics experiment was conducted to evaluate the motion and imaging ability of the designed endoscope device. In the planar bending experiment, the articulating module of the device was bent atop plotting paper, as depicted in Figure 9. Although the theoretical maximum bending angle was 177°, the maximum bending angle was set as 110° for structural safety of the articulating module, as depicted in Figure 9. The distance $d\alpha$ from the origin $O\alpha$ to the tip Ptip was 52 mm for a collinear position, which was defined as a bending angle of 0°, and $d\alpha$ was 45 mm for the maximum bending angle. The virtual arc for maximum bending can be expressed as $L\alpha=r\alpha\beta$, where $L\alpha$ is $d\alpha$ in the collinear position, $r\alpha$ represents the radius of the arc, and β represents the angle of the arc. The curvature κ is simply calculated via the inversion of $r\alpha$. In the experiment, β was approximately 0.698 rad (40.0°), $r\alpha$ was approximately 74.5 mm, and κ was approximately 0.013. These values can be adjusted by designing the joint angle between the two segments of the continuum body.

Figure 9. Experimental composite image showing the articulating range of the articulating module. The radius of curvature reaches 74.5 mm, and angle extends to 100°.



Meanwhile, the planar bidirectional bending for unit input movement of the motor was tested 25 times for each direction as shown in Figure 10; the resulting angles are plotted in Figure 11. In the experiment, the position of the tip of the continuum body was measured using an electromagnetic tracking system (Aurora System, Northern Digital Inc), and the setup is shown in Figure 10. Using the aforementioned smartphone app, the incremental input of the motor XL320 was varied by 8.9° per movement index. In Figure 11, the blue lines and gray areas denote averages and variances, respectively; further, a close trend of linear correlation is illustrated. The result demonstrate maximum standard deviations of 5.7° and 2.2° for right and left bending, respectively.

Figure 10. Sensor interface setup for the planar bending test.

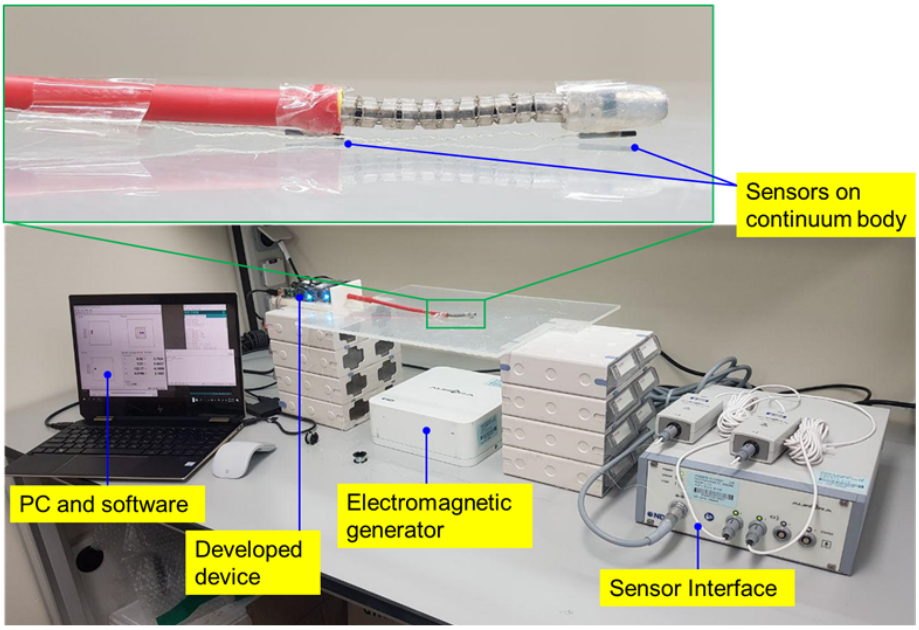
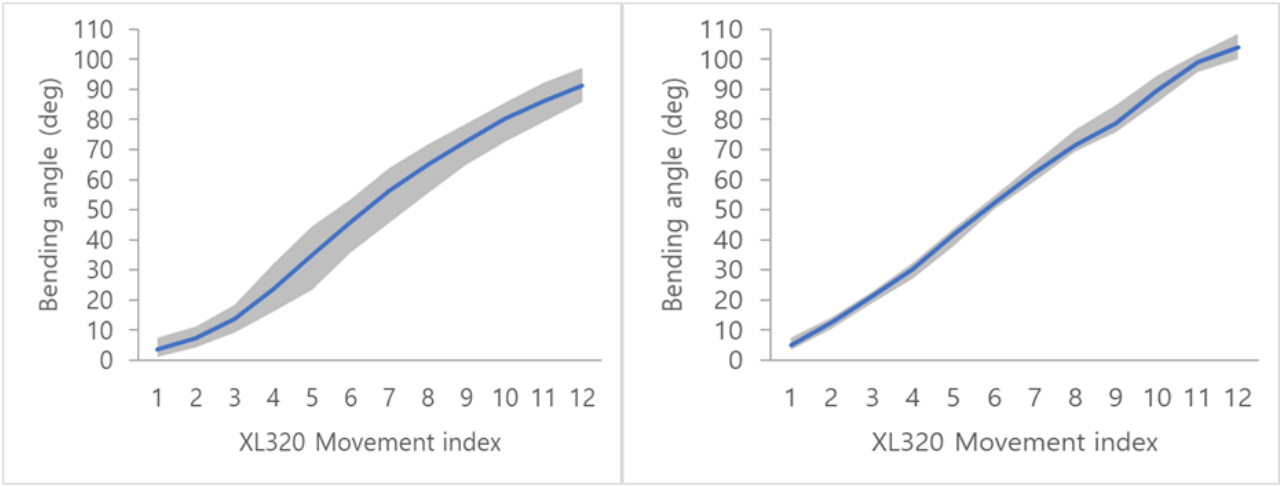
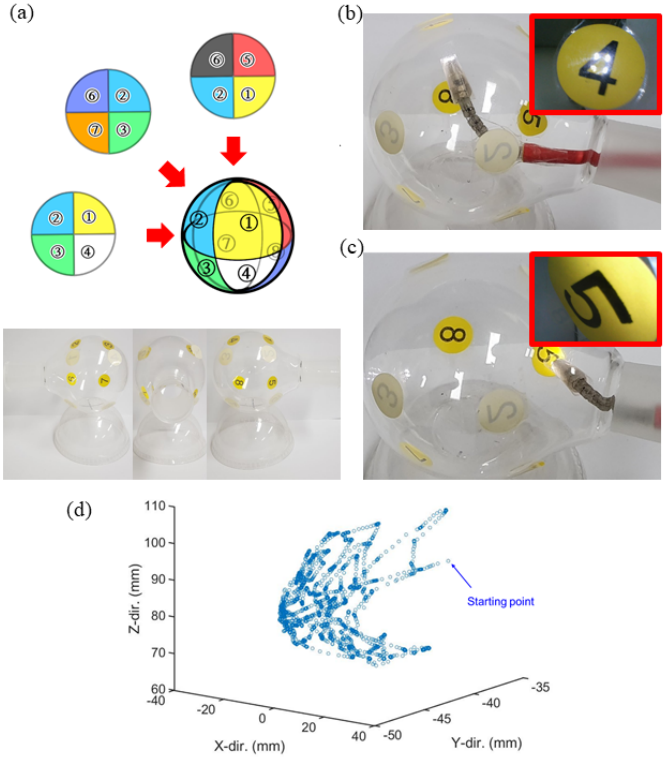


Figure 11. Relationship between the control input and the bending angle in the (a) right and (b) left directions.



In the spatial bending experiment, the spatial bending motion and image for randomly selected directions were verified by visualizing the insides of the transparent sphere octant. This surface (with a diameter of 85 mm) was prepared and divided into 8 sections, as depicted in Figure 12. The articulating module was made to face each divided section. The endoscope camera at the tip of the articulating module captured images of each octant and confirmed that the text on the sections was correctly recognized. Figures 12B and C show the position of the articulating module and images taken in sections ④ and ⑤. Figure 12D shows an example trajectory of the tip of the continuum body in a trial. In the trial, the tip moved from a starting point through section ⑦, middle section, section ③, middle section, and section ④, to middle section by a user's input of the smartphone app.

Figure 12. Spatial bending test of the continuum body: (a) test sphere divided into eight regions, (b) facing region ④ (inset: endoscopic view of region ④), (c) approaching region ⑤ (inset: endoscopic view of region ⑤), and (d) spatial trajectory of a trial.

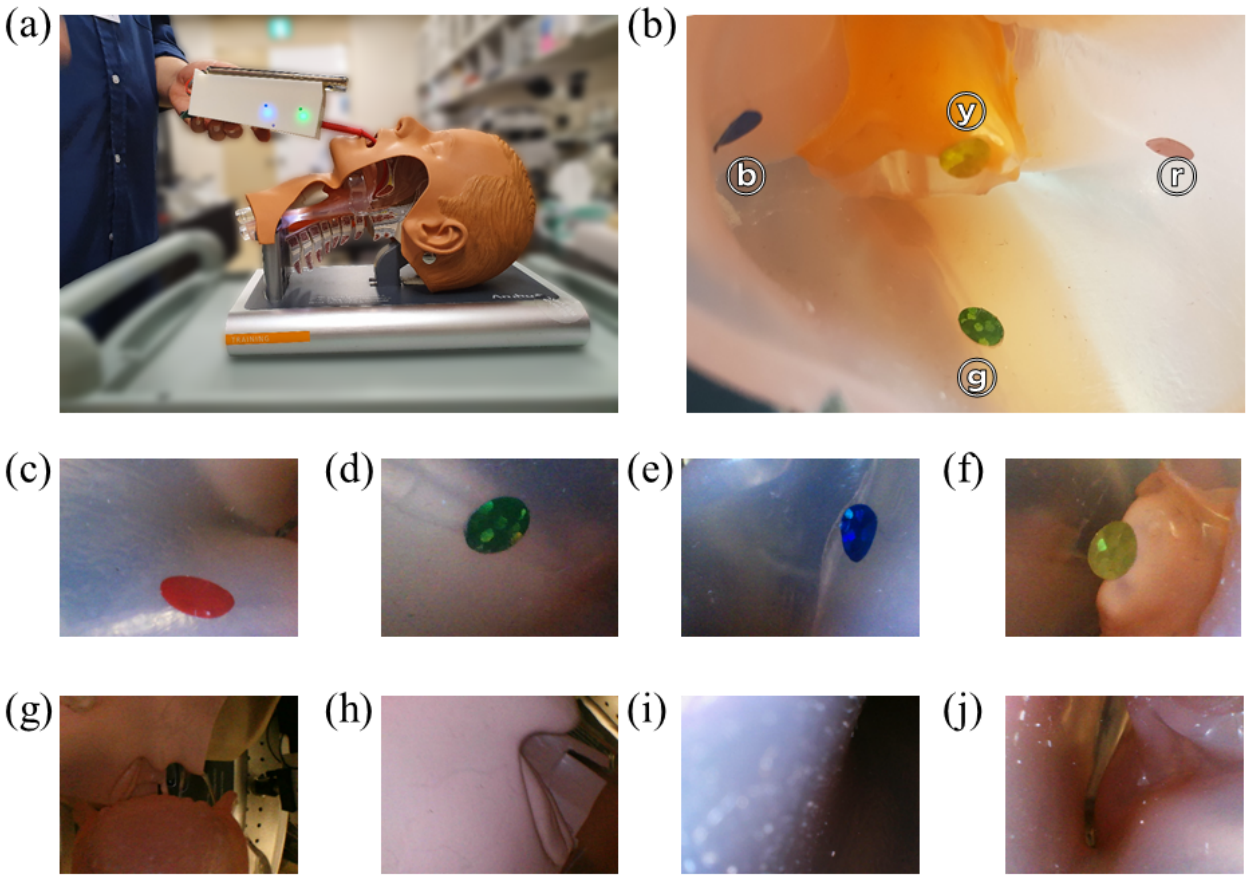


Phantom Imaging

To verify the ability to steer the articulating module, we placed 4 colored stickers (ventral, dorsal, left, and right) on the inner side of each of two human anatomical phantoms and captured images of the marked locations following a simulated clinical endoscope insertion. Figure 13A depicts the endoscope during the experiment. The operator was able to successfully insert the endoscope into a laryngopharyngeal phantom using only one hand. The placement of the colored marks is shown in Figure

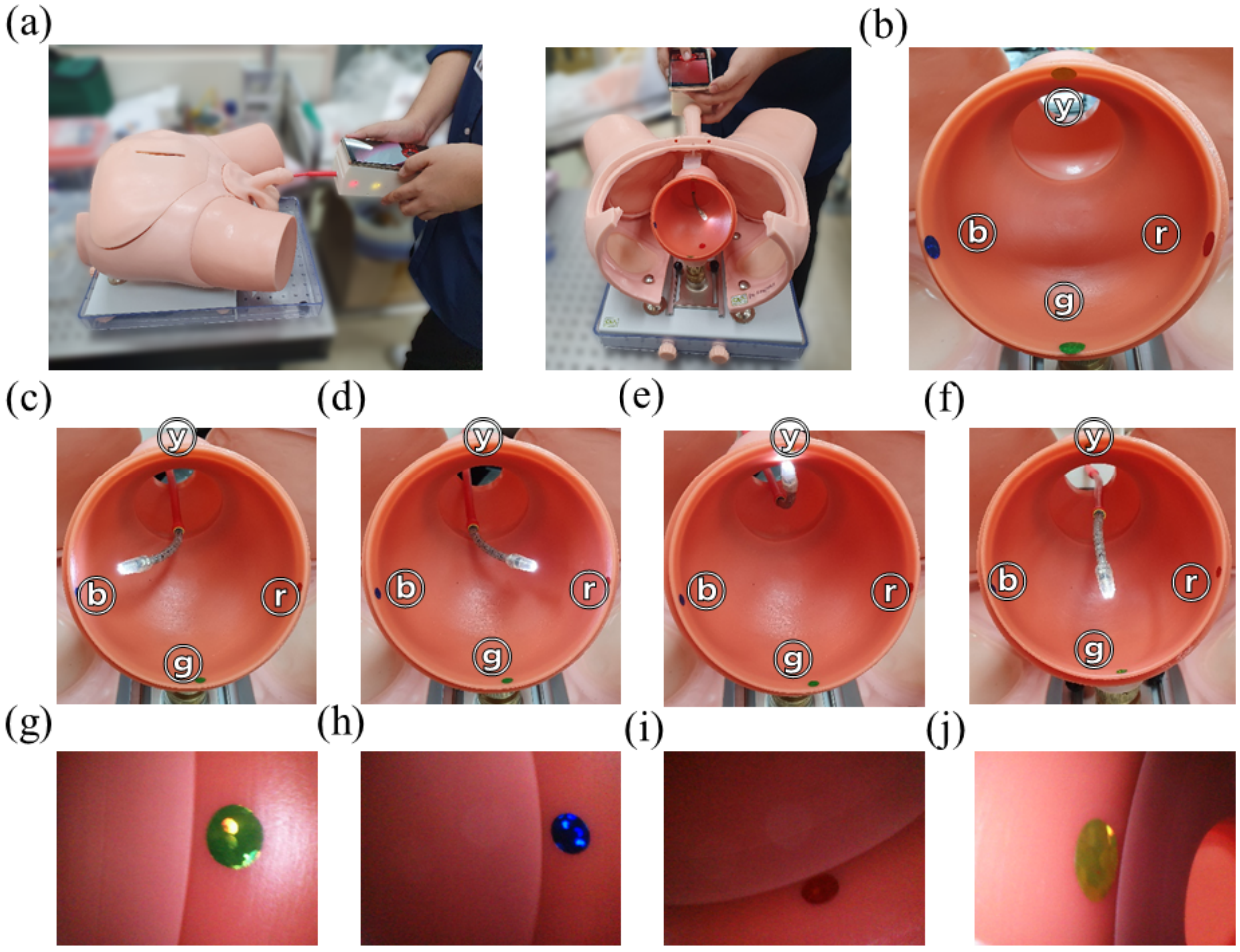
13B. The marks were in the oropharynx area: yellow, on the epiglottis; green, on the back wall of the oropharynx; and red and blue, on the right and left walls of the oropharynx, respectively. After the endoscope insertion, it was possible to image each of the marks by controlling the articulating module to face in the desired direction. Figures 13C-F present the resulting images. All 4 marks were clearly observed and within the focal range of the endoscope. Finally, we captured images while inserting the endoscope deeper into the phantom. Figures 13G-J show the insertion path through the laryngopharyngeal phantom.

Figure 13. Articulation experiment on the laryngopharyngeal phantom: (a) endoscope experiment in operation, (b) placement of stickers attached as indices in the oropharynx area of the phantom (upper side, yellow ⓨ; on the left, blue ⓑ; on the right, red ⓡ; and on the lower side, green ⓖ sticker), (c-f) images of the stickers obtained in the articulation mode, and (g-j) images captured by the camera, going deep from the oral cavity to the oropharynx.



A similar experiment simulated steering of the articulating module into and around the bladder. As before, 4 directions on the inner side of a bladder phantom were marked with colored adhesives. The endoscope was inserted into the bladder phantom via a simulated male urinary tract by an operator using one hand for stability and the other hand for insertion and steering, as shown in Figure 14A. The arrangement of the marks is shown in Figure 14B. The articulating module of the endoscope was able to face each marking successfully without movement at the proximal end of the endoscope module, as shown in Figures 14C-F. The successfully captured images of each of the four marks are shown in Figures 14G-J. The results, although captured on a phantom, indicate that the present rigid cystoscopy procedures can remain successful and be made significantly less invasive with the use of a similar articulating flexible endoscope design.

Figure 14. Articulation experiment on the bladder phantom: (a) endoscope experiment in operation, (b) placement of stickers on the inner side of the phantom (upper side, yellow ⓨ; on the left, blue ⓑ; on the right, red ⓡ; on the lower side, green ⓖ sticker, which was attached as an index), (c-f) articulation in four directions, and (g-j) corresponding images of the stickers obtained using the homebuilt smartphone app.



Discussion

Principal Findings

A portable wide-field endoscope combined with a smartphone was successfully developed. Using an app programmed in smartphone, an actively controlled articulable endoscope was demonstrated. Our device is compact in design, with a width and length similar to those of a smartphone. The endoscope system combined with smartphone can fit into one hand of either a male or a female operator. The weight of our device is 309 g without a smartphone. By attaching a smartphone, the weight can vary from 400 g to 500 g, depending on the conventional smartphone model. Therefore, as demonstrated experimentally, the user may freely conduct endoscopic procedures without being restricted to medical facilities such as clinics and hospitals. However, it is not limited to clinical applications but could be adopted for preclinical, veterinary, and industrial applications.

To facilitate the clinical use of the device, several human engineering factors were considered when implementing the UI design. The UI employs a joystick-like controller, and 4 fine control buttons are placed in the four corresponding directions, facilitating both intuitive and precise control, satisfying clinical needs. Placement of the additional buttons is low on the smartphone, allowing for ambidextrous control such that users may operate the virtual joystick, movie/still-image capture buttons, brightness adjustment, and contrast adjustment with ease even while focused on steering the endoscope.

The device runs on a smartphone, which is a highly available Linux computer, and the app for the device was developed in the C and Java languages, which are freely available. Because the application code is based on open-source projects, it can be easily modified by companies, local authorities, or other researchers for adding functionality. The Open Source Computer Vision Library (OpenCV) library is an open-source BSD-licensed library that includes hundreds of computer-vision algorithms [27] and provides programmatical brightness and contrast adjustment in app modules. Usually, brightness and contrast are the most significant factors that contribute to visual recognition; therefore, we made the contrast and brightness user-adjustable by adding postprocessing capabilities to the smartphone-based app. Because the app is based on OpenCV open source image processing program, complicated algorithms that can help diagnosis can be implemented. For example, OpenCV can provide the basic level of motion recognition, contour recognition, and background removal. Combined with the versatility and rapid processing ability of a smartphone, we can implement fairly complicated algorithms that generally require postprocessing on a desktop computer. Furthermore, the processed images are saved in the phone's internal memory

and can be easily accessed by other apps for further analysis or mobile sharing with other caregivers. This provides a new scheme of mobile examination and telemedicine. In industrial applications or at home, the user can easily send pictures to experts to examine data. In medical applications, endoscopes can be operated by trained users outside the hospital, and images can be shared with physicians and specialists for examination in real time. In addition, it enables easy dissemination of endoscopy devices in areas where medical benefits are insufficient.

The articulating module of the endoscope can be articulated in 2 DOFs. The actively controlled vertical and horizontal articulation of the articulating module make it easier to navigate through complex turns compared with conventional flexible endoscopes. This allows the user to easily find the desired location in a narrow area. To examine the feasibility of applying our articulating endoscope to medical examination, we used laryngopharyngeal and bladder phantoms. The pharynx and larynx are connected to the oropharyngeal area, which is distinguished by the epiglottis. The oropharynx and nasopharynx regions are smoothly connected to the pharynx but curve sharply when entering the larynx. Therefore, it is more difficult for the user to access the larynx when nonarticulating endoscopes are employed. As mentioned in the Introduction, patients with strong gag reflexes or limited jaw and neck movements are more likely to reject endoscopes when the endoscope touches the wall of the laryngopharyngeal complex. By using articulable endoscopes, the contact of endoscope tips with sensitive regions of the inner wall can be reduced. The depth this endoscopic device can reach depends on the length of each body shown in Figure 3. Each part in the endoscopic module body has the following dimensions: length of 52 mm for the distal, 150 mm for the flexible section, and 52 mm for the rigid section. Thus, the total length is approximately 254 mm. The average length of an adult's vocal cords, from mouth to vocal cords, is approximately 160 mm according to Cherng et al [28]. Therefore, the length of the developed device is sufficient for laryngopharyngeal applications. Furthermore, the length of conventional rigid resectoscopes for the transurethral resection of bladder tumors is 150 mm; therefore, the developed device can be employed for bladder diagnostic applications.

In addition to improving endoscopic access, the movability of the distal end of the endoscope significantly increases the range of visibility. For example, it is necessary to move the proximal part of both rigid and flexible cystoscopes if imaging on the opposite face of the bladder is desired. This movement can cause problems, including patient discomfort and destruction of the urethra. With the articulation, the endoscope unit attached to the distal articulating module can be articulated to face the desired direction without moving the proximal part of endoscope. The transparent sphere and bladder phantom experiments demonstrated this advantage of integrating the continuum body with the smartphone-attached endoscope. In the experiments, we succeeded in imaging all octants of the transparent sphere and all 4 marks on the bladder phantom without moving the proximal rigid part of the endoscope module. Thus, our device can face the desired part of the inner side of the bladder and obtain images [11].

The potential of portable smartphone-integrated medical devices involving endoscopes attached to the built-in camera of a smartphone has been discussed [29,30]. Using both rigid and flexible endoscopes, images are successfully captured [29]. Although their endoscope system is easy to manufacture, portable, and cheaper than conventional one, the device proposed in this study has several advantages over the devices of Bae et al [29]. First, our system uses an USB OTG connection, which allows the user to access external devices such as higher functions of cameras, data reader, or tablet computer, etc. This frees the endoscope from the limitations of internal smartphone cameras and from the other optical adjustments needed for magnification and precise alignment of the optical path. Moreover, since the performance of the external camera is independent from the specifications of the smartphone, as long as the device supports OTG connection, the external camera provides stable performance. Second, the proposed device is articulable and actively controlled, and thus allows imaging in deeper and from more complex organ structures compared with nonarticulable devices. This means that images of deeper positions can be obtained in a safe and stable state. Besides, when combined with biopsy devices, the procedures for biopsies and diagnosis in more difficult locations are easier. Third, depending on the capabilities of the mobile apps, it can be extended to a variety of application software such as image processing application, artificial intelligence video analysis application, statistical extraction application, and big data analysis application. By linking together, it is possible to add more and different technologies, thus overcoming the technical limitations of existing endoscopes.

Battery-based systems help users move freely out of a fixed area, independent from power generators. The proposed articulable device attached to a smartphone can be operated by a single operator outside the hospital and can be powered entirely by its own battery, isolated from the electrical grid. Since the battery is independent from the power generator, if the power runs out, it can sustain the system. The mobility of the device allows the user to employ the endoscope outside well-developed areas and makes it more suitable for use in underdeveloped countries, where electrical supplies are more likely to be interrupted due to power shortages, political complications, or poverty [31,32]. Moreover, portability may allow care providers to bring health services to rural areas and perform simple endoscopy for patients in underdeveloped areas. The smartphone allows visiting doctors or physicians to perform simple endoscopy, send images to faraway specialists for analysis, and then make appropriate treatment decisions.

The device is used for tubular organs and naturally opened orifices such as superficial layers of respiratory tract, urinary bladder, urethra, vaginal tract, and proximal part of colon to monitor lesion, polyps, stones, cancerous tissue, and abnormal dysplasia. Target users are medical professional, since nonprofessionals cannot keep up with medical knowledge or sterilization process, particularly in developing countries. Moreover, several countries have laws forbidding medical examination by nonprofessionals. Due to general medical guidelines, the endoscope probe must be sterilized before it is used and replaced every 20 rounds of examination session.

Finally, the portable articulable endoscope is far cheaper than a conventional endoscope system; the price to manufacture it is approximately US $480, excluding the smartphone. The estimated price will decrease with mass production. Table 2 compared available solutions that are being sold on the market. The price of nonarticulating endoscopes is considerably cheap; however, they are not articulable, have low image quality, and have no programmatic functions that help diagnosis. In contrast, conventional gastrointestinal endoscope systems have considerably more functions; however, they cannot implement complex algorithms such as machine learning or be taken out of the hospital facilities; further, they are very expensive. Therefore, as long as therapeutic accessories are not necessary, our system can be a simple and affordable solution for rapid diagnosis in developing countries.

Table 2. Comparison of available options.

Options	Smartphone-connected nonarticulating endoscope	Smartphone-based articulable endoscope, (proposed system)	Conventional gastrointestinal endoscope
Price (w/o smartphone), US$	10-100	473	20,000-120,000
Price (w/ smartphone), US$	300-1100 depending on smartphone model	700-1500 depending on smartphone model	N/A[a]
Probe tip DOF[b]	none	2 DOF	2 DOF

[a]N/A: not applicable.

[b]DOF: degree of freedom.

Considering that conventional endoscope systems from major companies and including camera, light source, etc, are usually priced above US $50,000 [29,33], the proposed device reduces the cost of care and increases access for people in underfunded and underdeveloped medical systems. With the acceptance of medical practitioners and up-to-date image processing algorithm, we could confirm that the proposed system has ample potential for use as a point-of-care diagnosis in developing countries.

Limitations

Since the developed system is a handheld type, movement of the endoscope probe may occur somewhat compared with the conventional mounted endoscope system. There is also a risk of sterilization and replacement of the endoscope probe, and there is a risk of breaking the drive wire if tension is applied that exceeds the operating range of the device. However, these sterilization, cleaning and wire breakage issues can be solved through simple replacement by making the endoscope probe into a module.

Conclusion

We designed, fabricated, and tested a portable smartphone-based and actively controlled articulable endoscope system. Microprocessors and smartphones have advanced to the point where they can serve as small computers. These portable devices open a new possibility for portable, cost-effective, and easy-to-operate at-home medical devices. In this study, we introduced a low-cost articulable endoscope system for mobile health care with an external camera as image sensor and smartphone as controller, providing advantages such as articulation, image control, mobility, and communication. We evaluated the feasibility of this system via phantom imaging. The results indicated that the device and homebuilt app have ample potential for preclinical, clinical, veterinary, and industrial applications. Specifically, the proposed endoscope could be a powerful tool for health care providers in regions with limited health care service.

Acknowledgments

This work was supported by the Basic Science Research Program (Grant No. 2019R1A2C2084122) and MRC grant (2018R1A5A2020732) through the National Research Foundation of Korea funded by the Ministry of Science & ICT; by the Ministry of Trade, Industry, and Energy under the Industrial Technology Innovation Program (10080726, 20000843); and by the Ministry of Health and Welfare, Republic of Korea (HI18C2391, HI17C2410).

Conflicts of Interest

None declared.

References

1. Davila RE, Rajan E, Adler D, Hirota WK, Jacobson BC, Leighton JA, ASGE. ASGE guideline: the role of endoscopy in the diagnosis, staging, and management of colorectal cancer. Gastrointest Endosc 2005 Jan;61(1):1-7. [doi: 10.1016/s0016-5107(04)02391-0] [Medline: 15672048]
2. Shah J. Endoscopy through the ages. BJU Int 2002 May;89(7):645-652 [FREE Full text] [doi: 10.1046/j.1464-410x.2002.02726.x] [Medline: 11966619]

3. Chuks N. Challenges of gastrointestinal endoscopy in resource-poor countries. In: Pascu O, editor. Gastrointestinal Endoscopy. London: IntechOpen; 2011.
4. del Rosario MB, Redmond SJ, Lovell NH. Tracking the evolution of smartphone sensing for monitoring human movement. Sensors (Basel) 2015 Jul 31;15(8):18901-18933 [FREE Full text] [doi: 10.3390/s150818901] [Medline: 26263998]
5. Yang K, Peretz-Soroka H, Liu Y, Lin F. Novel developments in mobile sensing based on the integration of microfluidic devices and smartphones. Lab Chip 2016 Mar 21;16(6):943-958 [FREE Full text] [doi: 10.1039/c5lc01524c] [Medline: 26899264]
6. Vadhanan P, Balakrishnan K. Evaluation of a low-cost videolaryngoscope: a randomized controlled pilot study. Anaesth Pain & Intensive Care. 4: 21; 2017. URL: https://www.apicareonline.com/index.php/APIC/article/view/84/847 [accessed 2020-08-10]
7. Amazon smartphone endoscope camera. URL: https://www.amazon.com/s?k=smartphone+endoscope+camera&s=price-asc-rank&page=4&qid=1572508774&ref=sr_pg_4 [accessed 2019-10-31]
8. Clayman RV, Reddy P, Lange PH. Flexible fiberoptic and rigid-rod lens endoscopy of the lower urinary tract: a prospective controlled comparison. J Urol 1984 Apr;131(4):715-716. [doi: 10.1016/s0022-5347(17)50594-7]
9. Gmeiner D, von Rahden BHA, Meco C, Hutter J, Oberascher G, Stein HJ. Flexible versus rigid endoscopy for treatment of foreign body impaction in the esophagus. Surg Endosc 2007 Nov;21(11):2026-2029. [doi: 10.1007/s00464-007-9252-6] [Medline: 17393244]
10. Eller R, Ginsburg M, Lurie D, Heman-Ackah Y, Lyons K, Sataloff R. Flexible laryngoscopy: a comparison of fiber optic and distal chip technologies. part 1: vocal fold masses. J Voice 2008 Nov;22(6):746-750. [doi: 10.1016/j.jvoice.2007.04.003]
11. Seklehner S, Remzi M, Fajkovic H, Saratlija-Novakovic Z, Skopek M, Resch I, et al. Prospective multi-institutional study analyzing pain perception of flexible and rigid cystoscopy in men. Urology 2015 Apr;85(4):737-741. [doi: 10.1016/j.urology.2015.01.007]
12. Krajewski W, Zdrojowy R, Wojciechowska J, Kościelska K, Dembowski J, Matuszewski M, et al. Patient comfort during flexible and rigid cystourethroscopy. Wideochir Inne Tech Maloinwazyjne 2016;2:94-97. [doi: 10.5114/wiitm.2016.60665]
13. Bagley DH. Intrarenal access with the flexible ureteropyeloscope: effects of active and passive tip deflection. J Endourol 1993 Jun;7(3):221-224. [doi: 10.1089/end.1993.7.221]
14. Wilking N, Petrelli NJ, Herrera-Ornelas L, Walsh D, Mittelman A. A comparison of the 25-cm rigid proctosigmoidoscope with the 65-cm flexible endoscope in the screening of patients for colorectal carcinoma. Cancer 1986 Feb 01;57(3):669-671. [doi: 10.1002/1097-0142(19860201)57:3<669::aid-cncr2820570345>3.0.co;2-w]
15. Nausheen F, Niknafs NP, MacLean DJ, Olvera DJ, Wolfe AC, Pennington TW, et al. The HEAVEN criteria predict laryngoscopic view and intubation success for both direct and video laryngoscopy: a cohort analysis. Scand J Trauma Resusc Emerg Med 2019 Apr 24;27(1):50 [FREE Full text] [doi: 10.1186/s13049-019-0614-6] [Medline: 31018857]
16. Baek MJ, Park S, Kim KH, Kim YH, Kim WK, Sun HY, et al. National trend of uroflowmetry, urodynamic study and cystoscopy considering the change in the population structure in Korea from 2010 to 2015. J Korean Med Sci 2018 May 14;33(20):e145 [FREE Full text] [doi: 10.3346/jkms.2018.33.e145] [Medline: 29760605]
17. Koeman K. Understanding USB On-The-Go, EDN. URL: https://www.edn.com/design/communications-networking/4346710/Understanding-USB-On-The-Go [accessed 2019-07-12]
18. Arduino Introduction. URL: https://www.arduino.cc/en/Guide/Introduction [accessed 2019-07-12]
19. DYNAMIXEL XL-320. Robotis. URL: http://www.robotis.us/dynamixel-xl-320/ [accessed 2019-07-12]
20. Burgner-Kahrs J, Rucker DC, Choset H. Continuum robots for medical applications: a survey. IEEE Trans Robot 2015 Dec;31(6):1261-1280. [doi: 10.1109/TRO.2015.2489500]
21. Kato T, Okumura I, Song S, Golby AJ, Hata N. Tendon-driven continuum robot for endoscopic surgery: preclinical development and validation of a tension propagation model. IEEE ASME Trans Mechatron 2015 Oct;20(5):2252-2263 [FREE Full text] [doi: 10.1109/TMECH.2014.2372635] [Medline: 26380544]
22. Camarillo DB, Milne CF, Carlson CR, Zinn MR, Salisbury JK. Mechanics modeling of tendon-driven continuum manipulators. IEEE Trans Robot 2008 Dec;24(6):1262-1273. [doi: 10.1109/tro.2008.2002311]
23. Jones B, Walker I. Kinematics for multisection continuum robots. IEEE Trans Robot 2006 Feb;22(1):43-55. [doi: 10.1109/TRO.2005.861458]
24. saki4510t/UVCCamera. GitHub. URL: https://github.com/saki4510t/OpenCVwithUVC [accessed 2019-07-12]
25. felHR85/UsbSerial. GitHub. URL: https://github.com/felHR85/UsbSerial [accessed 2019-07-12]
26. controlwear/virtual-joystick-android. GitHub. URL: https://github.com/controlwear/virtual-joystick-android [accessed 2019-07-12]
27. OpenCV: Introduction. URL: https://docs.opencv.org/master/d1/dfb/intro.html [accessed 2019-09-06]
28. Cherng C, Wong C, Hsu C, Ho S. Airway length in adults: estimation of the optimal endotracheal tube length for orotracheal intubation. J Clin Anesth 2002 Jun;14(4):271-274. [doi: 10.1016/s0952-8180(02)00355-0] [Medline: 12088810]
29. Bae JK, Vavilin A, You JS, Kim H, Ryu SY, Jang JH, et al. Smartphone-based endoscope system for advanced point-of-care diagnostics: feasibility study. JMIR Mhealth Uhealth 2017 Jul 27;5(7):e99 [FREE Full text] [doi: 10.2196/mhealth.7232] [Medline: 28751302]

30. Olawale F, Garwe D. Obstacles to the growth of new SMEs in South Africa: a principal component analysis approach. Afr J Bus Manag 2010;4(5):729-738.
31. Wu C, Wu S, Chen P, Lin Y. An innovative smartphone-based otorhinoendoscope and its application in mobile health and teleotolaryngology. J Med Internet Res 2014;16(3):e71 [FREE Full text] [doi: 10.2196/jmir.2959] [Medline: 24590187]
32. Olapade-Olaopa EO, Onawola KA. Challenges for urology in sub-Saharan Africa in 2006. J Men's Health Gender 2006 Mar;3(1):109-116. [doi: 10.1016/j.jmhg.2006.01.004]
33. Ciocîrlan. Low-cost disposable endoscope: pros and cons. Endosc Int Open 2019 Sep;7(9):E1184-E1186 [FREE Full text] [doi: 10.1055/a-0959-6003] [Medline: 31479502]

Abbreviations

3D: three-dimensional
CMOS: complementary metal-oxide semiconductor
DOF: degree of freedom
Open CV: Open Source Computer Vision Library
OTG: on-the-go
UI: user interface

Edited by G Eysenbach; submitted 14.11.19; peer-reviewed by C Poon, J Seitz; comments to author 27.01.20; revised version received 22.03.20; accepted 15.06.20; published 10.09.20.

Please cite as:
Moon Y, Oh J, Hyun J, Kim Y, Choi J, Namgoong J, Kim JK
Cost-Effective Smartphone-Based Articulable Endoscope Systems for Developing Countries: Instrument Validation Study
JMIR Mhealth Uhealth 2020;8(9):e17057
URL: http://mhealth.jmir.org/2020/9/e17057/
doi:10.2196/17057
PMID:32909951

Original Paper

Social and Behavior Change Communication Interventions Delivered Face-to-Face and by a Mobile Phone to Strengthen Vaccination Uptake and Improve Child Health in Rural India: Randomized Pilot Study

Mira Johri[1,2], MPH, PhD; Dinesh Chandra[3], MA; Karna Georges Kone[1,4], PhD; Marie-Pierre Sylvestre[1,5], PhD; Alok K Mathur[6], PhD, MBA; Sam Harper[7,8], PhD; Arijit Nandi[7,9], PhD

[1]Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montréal, QC, Canada

[2]Département de gestion, d'évaluation, et de politique de santé, École de santé publique de l'Université de Montréal, Montréal, QC, Canada

[3]Independent consultant, New Delhi, India

[4]Principal Technical Advisor for Health Financing at Management Sciences for Health (MSH)/USAID, Port-au-Prince, Haiti

[5]Département de médecine sociale et préventive, École de santé publique de l'Université de Montréal, Montréal, QC, Canada

[6]Indian Institute of Health Management Research University, Jaipur, India

[7]Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, QC, Canada

[8]Department of Public Health, Erasmus Medical Centre, Rotterdam, Netherlands

[9]The Institute for Health and Social Policy, McGill University, Montreal, QC, Canada

Corresponding Author:
Mira Johri, MPH, PhD
Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM)
850, rue St-Denis
Montréal, QC, H2X0A9
Canada
Phone: 1 514 343 7318
Email: mira.johri@umontreal.ca

Abstract

Background: In resource-poor settings, lack of awareness and low demand for services constitute important barriers to expanding the coverage of effective interventions. In India, childhood immunization is a priority health strategy with suboptimal uptake.

Objective: To assess study feasibility and key implementation outcomes for the Tika Vaani model, a new approach to educate and empower beneficiaries to improve immunization and child health.

Methods: A cluster-randomized pilot trial with a 1:1 allocation ratio was conducted in rural Uttar Pradesh, India, from January to September 2018. Villages were randomly assigned to either the intervention or control group. In each participating village, surveyors conducted a complete enumeration to identify eligible households and requested participation before randomization. Interventions were designed through formative research using a social marketing approach and delivered over 3 months using strategies adapted to disadvantaged populations: (1) mobile health (mHealth): entertaining educational audio capsules (edutainment) and voice immunization reminders via mobile phone and (2) face-to-face: community mobilization activities, including 3 small group meetings offered to each participant. The control group received usual services. The main outcomes were prespecified criteria for feasibility of the main study (recruitment, randomization, retention, contamination, and adoption). Secondary endpoints tested equity of coverage and changes in intermediate outcomes. Statistical methods included descriptive statistics to assess feasibility, penalized logistic regression and ordered logistic regression to assess coverage, and generalized estimating equation models to assess changes in intermediate outcomes.

Results: All villages consented to participate. Gaps in administrative data hampered recruitment; 14.0% (79/565) of recorded households were nonresident. Only 1.4% (8/565) of households did not consent. A total of 387 households (184 intervention and 203 control) with children aged 0 to 12 months in 26 villages (13 intervention and 13 control) were included and randomized. The end line survey occurred during the flood season; 17.6% (68/387) of the households were absent. Contamination was less than 1%. Participation in one or more interventions was 94.0% (173/184), 78.3% (144/184) for the face-to-face strategy, and 67.4% (124/184) for the mHealth strategy. Determinants including place of residence, mobile phone access, education, and female

empowerment shaped intervention use; factors operated differently for face-to-face and mHealth strategies. For 11 of 13 intermediate outcomes, regression results showed significantly higher basic health knowledge among the intervention group, supporting hypothesized causal mechanisms.

Conclusions: A future trial of a new intervention model is feasible. The interventions could strengthen the delivery of immunization and universal primary health care. Social and behavior change communication via mobile phones proved viable and contributed to standardization and scalability. Face-to-face interactions remain necessary to achieve equity and reach, suggesting the need for ongoing health system strengthening to accompany the introduction of communication technologies.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 44840759; https://doi.org/10.1186/ISRCTN44840759

(JMIR Mhealth Uhealth 2020;8(9):e20356) doi:10.2196/20356

KEYWORDS

randomized controlled trial; immunization programs; child health; mHealth; health promotion; health services accessibility; implementation science; pilot projects; developing countries; global health

Introduction

Background and Rationale

Expanding coverage of effective interventions is a critical challenge for many low- and middle-income countries (LMICs). In addition to technical improvements in service delivery, improving coverage often hinges critically on enhanced awareness and demand for services on the part of beneficiaries. Furthermore, in settings of low literacy, deep poverty, and poor access to information, behavior change is extremely challenging.

Immunization is a priority health strategy for LMIC policy makers seeking to advance the 2030 United Nations (UN) Sustainable Development Goals (SDGs) due to its inherent value in reducing the burden of disease and its potential role as a lever for health system strengthening. Immunization reaches more households than any other health service, bringing communities into regular contact with the health system [1]. The immunization platform can potentially be used to strengthen the delivery of universal primary health care, universal health coverage, and meet other SDG targets [1]. This approach may be particularly salient in areas where vaccination delivery systems function reliably, but important gaps exist in the delivery of other health services. In these contexts, increasing immunization coverage offers a potential pathway to expand the range and reach of health services and to advance a holistic health agenda.

In India, the government has prioritized immunization, making remarkable gains in recent years. However, coverage continues to fall short of the target to fully immunize 90% of India's infants against 7 vaccine-preventable diseases by 2020 [2]. Nevertheless, immunization delivery now outperforms other services, offering a potential lever for system strengthening. In rural north India, for example, research shows that high-priority primary care interventions, including vaccination, are being delivered quite well, whereas other basic health promotion and prevention services are largely not provided, constituting a critical missed opportunity for population health [3]. On the basis of analysis of Indian immunization program data, achieving and sustaining vaccine coverage targets especially requires new strategies to address gaps in beneficiary demand [2]. Recent systematic reviews and meta-analyses have demonstrated that knowledge translation and education strategies, such as those offering education at village meetings or at home, are likely to improve vaccination coverage [4,5]. However, strategies based on face-to-face communication may be challenging to standardize and deliver at scale.

The widespread availability of mobile phones in LMICs has stimulated interest in the potential of mobile health (mHealth) interventions to achieve health objectives. A recent systematic review found that mHealth interventions can improve maternal and neonatal service delivery and that text-based vaccination reminders are associated with improved vaccination coverage [6]. Although their potential for scalability at low cost is attractive, whether mHealth interventions can be effective for highly disadvantaged populations facing substantial barriers due to poverty, low literacy, and gender norms is uncertain.

Goal of This Study

We developed the *Tika Vaani* (*vaccine voice* in Hindi) model to educate beneficiaries about immunization and basic child health themes, dispel misinformation, and empower households to better care for their children and themselves. A key distinguishing feature of the model is that it integrates an mHealth component to increase standardization and scalability of social and behavior change communication. The interventions were delivered through small face-to-face meetings and via mobile phone using strategies (context-appropriate audio messages delivered via automated phone calls) adapted to disadvantaged populations. We conducted a pilot randomized controlled trial (RCT) applying an implementation research lens to inform a future large-scale RCT. This paper presents the main (quantitative) evaluation of the pilot trial focusing on two objectives: (1) to assess the feasibility of processes critical to the success of the main study (recruitment, randomization, retention, and contamination) and intervention uptake (adoption) and (2) to study key implementation outcomes to optimize successful delivery of the interventions at scale [7]. Objectives pertained to cluster and individual levels. A companion paper presents findings related to intervention fidelity, acceptability, and appropriateness using mixed (qualitative and quantitative) methods (Pérez et al, unpublished data, 2020).

Methods

Trial Design

In keeping with the plan for the main study, this pilot study adopted a cluster-randomized design with a 1:1 allocation ratio. A cluster design was chosen owing to the nature of the study interventions: face-to-face interventions are structured around communities rather than individuals, whereas mHealth interventions have a possibility of contamination. Clusters were rural villages in a district of Uttar Pradesh (UP), India. Villages were randomly assigned to either the intervention or control group (CG). The protocol was registered in a WHO International Clinical Trials Registry Platform-compliant registry (ISRCTN44840759 doi.org/10.1186/ISRCTN44840759). There were no important changes to methods after trial commencement. We originally sought to register the trial in the Clinical Trials Registry–India (CTRI), which is free of charge and has as the mission to enroll all clinical trials conducted in India. The CTRI took several months to follow-up; in the interim, we applied to a different registry. Owing to the delay caused by waiting and changing registries, the trial was registered shortly after patient enrollment was completed.

Participants

Setting and Location

India's most populous state of over 200 million residents, UP is an area of national focus due to lagging health and development indicators. Hardoi (population 4 million; under-5 mortality rate 118 per 1000; cf. UP under-5 mortality rate 90 per 1000, India under-5 mortality rate 57.3 per 1000) [8,9]) is a low-performing, rural district within UP comprising 19 administrative blocks. Thanks to recent Government of India (GoI) initiatives, the proportion of fully immunized children aged between 12 and 23 months in Hardoi district rose from 26.5% in 2007-2008 [10] to an estimated 65.9% (95% CI 62.0%-69.8%) in 2018 when this study was conducted [2]. A single administrative block of the Hardoi district was selected for this pilot based on criteria reflecting logistics and needs.

Eligibility Criteria

Villages (clusters) were eligible for inclusion if they had less than 4000 inhabitants and were located in Bawan Block, Hardoi, UP. In participating villages, interventions were offered to all residents. Participants in the baseline survey were consenting primary caregivers of children aged between 0 and 12 months residing in a study village. We excluded those who were not able to understand and speak Hindi or Urdu or those who did not intend to reside in the village for the study duration (6 months). The same individuals were approached for the end line survey.

The survey sampling unit was the household. We conducted a door-to-door census of the village and cross-checked administrative records from the Anganwadi workers (AWW) and Accredited Social Health Activists (ASHA) to identify all households containing a child in the age range of 0 to 12 months within each village. These households constituted our primary target group.

Interventions

Formative Research

From January 1, 2017, to January 10, 2018, we conducted formative research using a social marketing approach to inform intervention design [11]. An iterative, participatory approach involving cocreation was favored to validate the need for the interventions, to make the interventions more compelling and linguistically and culturally appropriate, and to tailor approaches to different user segments [11]. Content was designed by Gram Vaani, an Indian social enterprise specializing in community media platforms for low-literacy rural populations, and Jagriti, a local NGO. Content fostered equity, empowerment, and social inclusion through positive portrayals of diverse characters. Technical experts assured information quality, including members of the research team and India's Ministry of Health and Family Welfare [11]. Extensive adaptations to interventions were made to meet target population needs during the formative research phase [12]. During the pilot RCT, all intervention components were frozen for evaluation, and deviations to intervention fidelity were systematically monitored (Pérez et al, unpublished data, 2020.

Pilot RCT

The study interventions took place over a 3-month period and offered social and behavior change communication (SBCC) for members of the general public in rural Indian villages, addressing topics related to child health. The primary target group was families residing in a selected village with a child in the age range of 0 to 12 months. Although vaccination was the primary focus of the study, the SBCC interventions addressed additional areas stipulated to be co-delivered with immunization during Village Health and Nutrition Days (VHNDs), such as health education related to health care entitlements; prevention, recognition, and management of common infectious diseases (diarrhea, pneumonia, dengue, and chikungunya); nutrition; and water, sanitation, and hygiene (WASH).

SBCC materials were delivered through 2 channels: (1) mHealth: educational audio capsules in entertaining formats (edutainment) and voice reminders for immunization sessions broadcast via mobile phone and (2) face-to-face: community mobilization activities, consisting of 1 large introductory meeting to introduce the project to communities and 3 small meetings offered to each participant covering specific themes. For the mHealth component, push messages (automated dial outs) and voice-based reminders were privileged owing to low education level and comfort with technology. For the face-to-face component, small group meetings were held separately for men and women and in different geographical locations within villages to ensure ease of communication. mHealth vaccination reminders were based on the child's birthdate and offered only to the target group; however, other interventions (mHealth edutainment and face-to-face meetings) were open to all village residents. Community workers (AWW and ASHAs) were encouraged to participate and received advance access to intervention materials. All interventions were free of charge to end users. The CG received standard GoI health and welfare services.

The mHealth strategy (Tika Vaani SBCC Version 1.5, released on July 7, 2017) was designed and delivered by Gram Vaani, a social tech startup incubated out of the Indian Institute of Technology Delhi, using the Mobile Vaani Interactive Voice Response System. Access was free and open to anyone who called the number. The participants could give a missed call to access the platform, and as a result receive a callback enabling them to access all content, record any queries or feedback, or be connected to a live expert. To simplify access, consenting households in intervention villages with children aged less than 12 months at baseline received automated outbound calls twice per week. In total, 26 content pieces were offered. In addition, child vaccination reminders were sent to target group households. Small group meetings lasting approximately 1 hour involving 2 trained facilitators with a minimum of 12 years of education were held once per month and open to all village residents. The access number was shared at each meeting. A video describing Mobile Vaani is available [13]. Additional information relevant to scale up and replication of the Tika Vaani system is available [11] and content is accessible [14]. A comprehensive intervention description is provided in Multimedia Appendix 1 [15]. The evaluation was conducted by a team of academic specialists distinct from the developers.

Outcomes

The pilot study considered a range of implementation outcomes (Table 1) [16].

Table 1. Outcome variables and data sources for the Tika Vaani social and behavior change communication pilot study.

Outcomes	Definition	Approach	Analysis sample	Data sources
Primary outcomes[a]				
Feasibility of the future main study	Ex-ante success criteria • Recruitment • Randomization • Retention • Contamination	Quant[b]	IG[c] and CG[d]	• Project records (all) • IVR[e] platform • HH[f] surveys (contamination)
Uptake (adoption)	Participation in • Small group meetings • mHealth[g]	Quant	IG	• Project records (meetings) • IVR platform (mHealth)
Secondary outcomes				
Acceptability and appropriateness	Perception among stakeholders that an intervention is agreeable, suitable, relevant, useful, and credible	Mixed methods[h]	IG	Refer to the study by Pérez et al (unpublished data, 2020)
Fidelity	Ability to deliver the interventions as planned	Mixed methods	IG (and CG)	Refer to the study by Pérez et al (unpublished data, 2020)
Coverage	The degree to which a population eligible to benefit from an intervention actually receives it	Quant	IG	• HH surveys • Project records (meetings) • IVR platform (mHealth)
Adequacy of the program theory	Intermediate outcomes reflecting changes in knowledge, attitudes, and practices of end users	Quant	IG and CG	HH surveys

[a]Outcomes and definitions adapted from the study by Peters et al [16].

[b]Quant: quantitative.

[c]IG: intervention group.

[d]CG: control group.

[e]IVR: interactive voice response.

[f]HH survey: household survey.

[g]mHealth: mobile health.

[h]Mixed methods: quantitative and qualitative.

Primary Outcomes

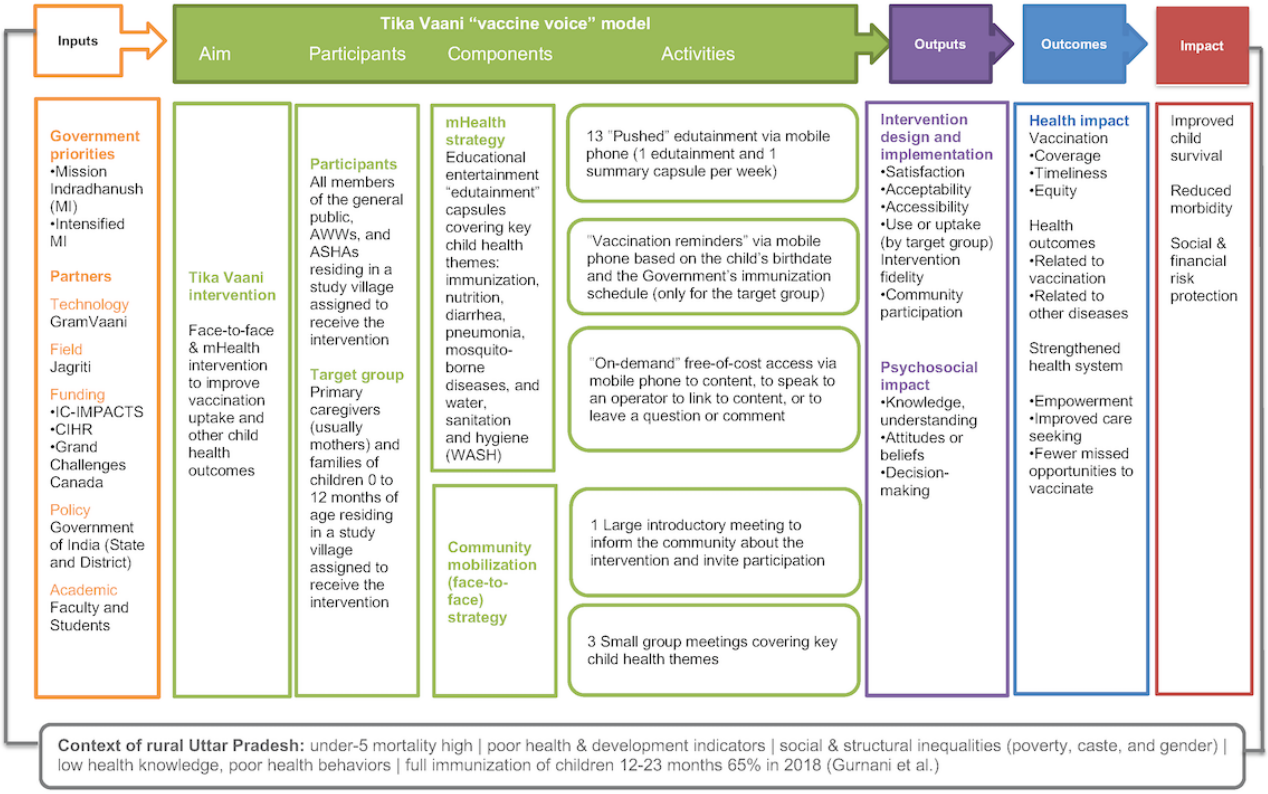
We established ex-ante criteria for the feasibility of the main study related to recruitment, randomization, retention, and contamination. We were concerned about contamination among treatment groups for mHealth services, as the phone number is easily shared. We viewed a contamination proportion exceeding 15% as a threat to the feasibility of adopting a cluster-randomized design with village as the unit of randomization and geographical distances between villages

(mean 15 km; range 5 km-50 km) similar to those in the pilot. As health interventions must achieve sufficient uptake to impact population health, we also established minimum criteria for participation in the new interventions.

Secondary Outcomes

We present quantitative findings for 2 secondary outcomes: (1) coverage or the extent to which the interventions reached specific populations and (2) adequacy of the program theory. We constructed a logic model describing the hypothesized program impact pathway and mechanisms of action (Figure 1) and adapted an established vaccination communication taxonomy to define indicators [17]. We compared treatment groups on *outputs* (intermediate outcomes, such as knowledge and attitudes) related to the intervention theory of change.

Figure 1. Conceptual model of the intervention. ASHAs: Accredited Social Health Activists; AWWs: Anganwadi Workers; CIHR: Canadian Institutes for Health Research; IC-IMPACTS: the India-Canada Centre for Innovative Multidisciplinary Partnerships to Accelerate Community Transformation and Sustainability; WASH: water, sanitation, and hygiene.



Data Sources

Quantitative data were collected from the following sources:

- Face-to-face surveys: Interviewer-administered household surveys were conducted in all participating study villages. Surveys were administered at baseline before random group assignment and approximately 5 months later following interventions at end line.
- Project administrative records: Standardized forms to assess delivery of study procedures and interventions were maintained by field staff to facilitate structured observation and data capture.
- Interactive voice response (IVR) system: The IVR system automatically recorded all calls to the platform. IVR data can be disaggregated by various fields including caller phone number, date, time, frequency, duration (seconds), call type, content type, and user characteristics. IVR data were linked to phone numbers provided by target households during the baseline survey to map calls sent and/or received.

Variables

We measured the use of the IVR through the number and duration of calls from a unique phone number. We considered that an mHealth item was *received* if the caller remained connected for 80% or more of the item duration. Contamination was defined as the proportion of calls originating from control villages and assessed using 2 data sources: (1) the IVR system to identify unknown numbers and trace nonregistered callers to identify call origin and (2) questions were included about intervention use in the end line survey administered to the treatment and CGs (Multimedia Appendix 1). To construct wealth quintiles, we performed principal component analysis to create a relative index of household wealth from a list of assets used in major household surveys [10,18]. We used this index to divide the sample into quintiles [19]. A similar approach was used to construct women's empowerment terciles. Caste was represented in 4 categories (general, other backward caste, scheduled caste, or scheduled tribe) ranging from most to least advantaged, as for Indian national surveys. By convention, caste categories are applied to all population groups, irrespective of religion. For intervention group (IG) households only, implementation teams assessed whether households were able

to attend small group meetings based on geographical distance from their place of residence to anticipated meeting sites. This categorization was made at the study baseline, before undertaking interventions.

Sample Size

Although the pilot was a cluster-randomized two-group study, the study size was based on the rate of contamination among controls. We estimated the required sample size needed for the CG using methods for a one-group descriptive study. We assumed that the true proportion of contamination (calls originating from controls) was 10%, that contamination was most likely to arise from parents of young children, and that there would be 20 households with children aged less than 12 months per village. On the basis of these inputs and using a binomial (*exact*) calculation, we would require 200 households in the CG to detect a 95% confidence interval of 6.2% to 15.0% [20]. The total sample size for the pilot was therefore set at double this number or 400 households.

Village Selection

The sampling frame was informed by the 2011 census [21], which indicated that the Bawan Block had a total population of approximately 234,000, including 217,000 rural residents distributed among 143 villages. We eliminated 3 urban villages, 11 villages with a population exceeding 4000, 15 villages recorded as having 0 population, and 57 villages in which we had previously worked (so as to gain experience in a treatment-naïve population). This left a sampling frame of 57 villages, from which an initial 20 villages were randomly selected using Microsoft Excel. The number of children in the target age range per village was not known in advance of the baseline survey. As villages vary in size, to attain our sample size target, we decided a priori that (1) any village with more than 1 but less than 10 children in the target age range would be retained and another randomly selected village would also be added and (2) villages with no children in the target age range would be dropped.

Randomization

Sequence Generation and Allocation Concealment

Villages were assigned to either intervention or control using simple randomization with a 1:1 allocation following a computer-generated randomization schedule. The random allocation sequence was generated at the Centre de recherche du Centre hospitalier de l'Université de Montréal by a professional statistician (MPS) using commands for random samples and permutations in the R environment for statistical computing [22] and kept in a password-protected computer. The statistician was not involved in study implementation. Before the release of the randomization code, only the statistician had access to the allocation sequence. Randomization code was released all at once, and treatment groups were assigned only after completing all recruitment procedures and baseline measurements.

Implementation

Field team leaders enrolled clusters by contacting village officials in person to explain study aims and activities and request consent to participate. Subsequently, in each participating village, surveyors conducted a complete enumeration to identify all households with children in the target age range and directly approached all such households to request participation in the baseline survey and pilot study. Consent was sought before randomization. No advertisements were used for recruitment, and no incentives or rewards were offered for participation. Surveyors communicated group assignments personally to households.

Blinding

Due to the nature of the interventions, neither participants nor those delivering interventions were blinded to the group assignment. We hired independent surveyors at end line to assess study outcomes. These surveyors were not informed about study aims or group assignments.

Statistical Methods

Descriptive Analyses

We used counts, frequencies, and proportions to summarize categorical data, and means and standard deviations for continuous variables. We assessed bivariate associations using univariable logistic regression for continuous variables and the chi-square test for categorical variables.

Multivariable Analyses

Coverage

We studied the degree to which target beneficiaries (IG households with a child aged less than 12 months at baseline) received the interventions.

To investigate patterns of uptake, we developed separate models for each intervention component: immunization reminders (mHealth), edutainment capsules (mHealth), and small group meetings (face-to-face). Outcomes were modeled as binary (0 uptake vs 1 or more instances of uptake). We followed guidance for the use of logistic regression in small data sets [23,24]. To identify candidate predictors of uptake and use, we developed a conceptual framework informed by the scientific literature and expert knowledge (Multimedia Appendix 1). The conceptual framework considered socioeconomic determinants, physical and infrastructure barriers, access barriers related to mobile phone use within households, and women's empowerment. Together, these variables represented 17 degrees of freedom. To develop the full models for implementation, we empirically refined the choice of variables to respect a minimum of roughly 10 events per variable for accurate prediction of binary outcomes [23]. Specifically, we excluded candidate variables if the bivariate chi-square test showed no relationship between predictor and outcome at the level of $P<.25$. We fit full statistical models using the Firth (penalized maximum likelihood) logistic regression to avoid overfitting [23] and handle data separation. The final models were fit within a cluster bootstrap algorithm (1000 iterations).

To study the determinants of intensity of participation, we repeated analyses specifying ordered logistic regression models for 2 outcomes: (1) the number of mHealth items heard and (2) the number of small group meetings attended. We tested the

proportional odds assumption using an approximate likelihood ratio test [25]. All models used robust standard errors to account for clustering.

Adequacy of the Program Theory

We studied intervention impact on intermediate outcomes using generalized estimating equations (GEE): (1) we used the differences-in-differences method to study changes in variables measuring immunization knowledge in the 2 study groups between the baseline survey and the end line survey using unadjusted regression coefficients (with their 95% CIs) for the interaction between group (intervention or control) and time period (end line or baseline) [26]. (2) To assess knowledge of other basic health topics (assessed only at study end line), we estimated the probability of correct responses at the end line among those receiving the intervention (vs controls). All GEE models estimated binary outcomes with an exchangeable correlation structure adjusted for village-level clustering and robust standard errors. We ran crude models and models adjusted for unbalanced variables following randomization and prespecified potential confounders.

Feasibility outcomes were assessed using the intention-to-treat (ITT) sample; no clusters and no participants were excluded. Analyses of intervention uptake and coverage used the ITT IG; no clusters and no participants randomized to the IG were excluded. To assess the program theory, we analyzed intermediate outcomes using the sample that participated at both baseline and end line, for which 0 clusters, 68 households, and 69 caregivers were lost to follow-up, which is equivalent to an observational sample. For 2 households, missing data on caste were imputed based on the locality of residence within the village. There were no other missing data. Analyses were conducted in Stata 15 (Stata Corporation).

Research Ethics and Informed Consent

Permission was granted by the Institutional Committee for Ethics and Review of Research, Indian Institute of Health Management Research, Jaipur, on January 10, 2017, and by the Comité d'éthique de la recherche du Centre hospitalier de l'Université de Montréal (Research Ethics Committee of the University of Montreal Hospital) on January 11, 2017 (Reference number 16.084). All participants provided written, in-person informed consent. After completing the study, we offered CG residents access to the mHealth interventions.

Results

Participants

The baseline survey and recruitment took place from January 19 to February 19, 2018. We approached 29 villages and 100% (29/29) consented to participate. Recruitment of individual participants was complicated by gaps in administrative data, as 13.9% (79/565) of recorded households were in fact nonresident. Only 1.4% (8/565) of the candidate households did not consent to participate. A total of 391 (185 *IG* and 206 *CG*) caregivers of children aged 0 to 12 months in 387 (184 IG and 203 CG) households in 26 (13 IG and 13 CG) villages were included in the study (Figure 2). Interventions were delivered from March 21 to July 9, 2018. The end line survey took place from July 17 to August 20, 2018, during the annual monsoon floods. Many households (17.6%, 68/387) were absent during the study end line; loss to follow-up was non-differential (31 IG and 37 CG). The trial ended when planned activities were successfully completed; Figure 2 describes the progress of participants through the trial.

Characteristics of the participating individuals (Table 2) and villages (Table 3) were similar across treatment groups at baseline. The CG was advantaged in terms of assets (wealth quintiles) and cell phone network quality.

Figure 2. Flow diagram of the parallel group cluster trial. ITT: intention-to-treat.

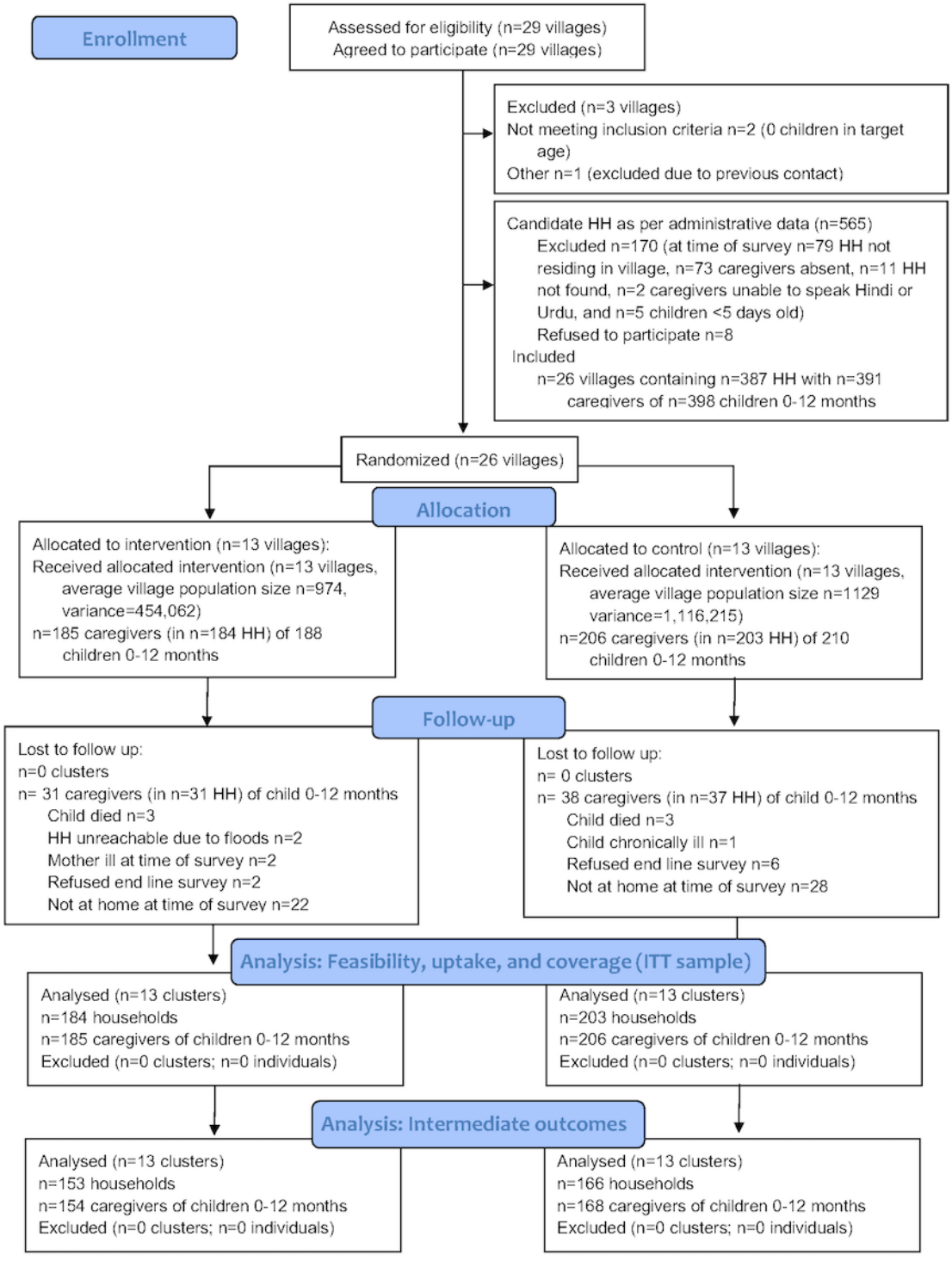


Table 2. Baseline characteristics of participating households, by treatment group.

Variable[a]	Intervention (n=184), n (%)	Control (n=203), n (%)	All participants (n=387), n (%)
Wealth index (quintile)			
(Q1) Lowest	48 (26.1)	30 (14.8)	78 (20.2)
(Q2)	37 (20.1)	40 (19.7)	77 (19.9)
(Q3)	37 (20.1)	41 (20.2)	78 (20.2)
(Q4)	28 (15.2)	49 (24.1)	77 (19.9)
(Q5) Highest	34 (18.5)	43 (21.2)	77 (19.9)
Religion[b]			
Hindu	181 (98.4)	176 (86.7)	357 (92.3)
Muslim	3 (1.6)	27 (13.3)	30 (7.8)
Caste[c]			
General	38 (20.7)	37 (18.2)	75 (19.4)
Other backward caste	89 (48.4)	80 (39.4)	169 (43.7)
Scheduled caste	57 (31.0)	86 (42.4)	143 (37.0)
Maternal education (years)			
None (0)	62 (33.7)	75 (37.0)	137 (35.4)
Primary (1-8)	85 (46.2)	84 (41.4)	169 (43.7)
Secondary (9-12) or more	37 (20.1)	44 (21.7)	81 (20.9)
Paternal education (years)			
None (0)	29 (15.8)	26 (12.8)	55 (14.2)
Primary (1-8)	90 (48.9)	99 (48.8)	189 (48.8)
Secondary (9-12) or more	65 (35.3)	78 (38.4)	143 (37.0)
HH lives far from meetings[d]			
No	142 (77.2)	N/A[e]	N/A
Yes	42 (22.8)	N/A	N/A
Cell phone network poor			
No	163 (88.6)	203 (100.0)	366 (94.6)
Yes	21 (11.4)	0 (0.0)	21 (5.4)
HH owns a mobile phone			
No	11 (6.0)	15 (7.4)	26 (6.7)
Yes	173 (94.0)	188 (92.6)	361 (93.3)
In this HH, mother has phone			
No	106 (57.6)	134 (66.0)	240 (62.0)
Yes	78 (42.4)	69 (34.0)	147 (38.0)
In this HH, father has phone			
No	68 (37.0)	76 (37.4)	144 (37.2)
Yes	116 (63.0)	127 (62.6)	243 (62.8)
In this HH, someone else has phone			
No	162 (88.0)	166 (81.8)	328 (84.8)
Yes	22 (12.0)	37 (18.2)	59 (15.3)
Mother's phone access			
Cannot access	90 (48.9)	88 (43.4)	179 (45.0)

Variable[a]	Intervention (n=184), n (%)	Control (n=203), n (%)	All participants (n=387), n (%)
Can use easily	94 (51.1)	115 (56.7)	209 (54.0)
Mother can dial phone			
No	47 (25.5)	55 (271)	102 (26.4)
Yes	137 (745)	148 (72.9)	285 (73.6)
Mother's frequency of phone use			
Rarely	27 (14.7)	30 (14.8)	57 (14.7)
When needed	102 (55.4)	106 (52.2)	208 (53.8)
Almost daily	55 (29.9)	67 (33.0)	122 (31.5)
Women's empowerment score (tercile)			
Lowest	83 (45.1)	94 (46.3)	177 (45.7)
Average	74 (40.2)	84 (41.4)	158 (40.8)
Highest	27 (14.7)	25 (12.3)	52 (13.4)
Permission to attend village meetings			
No	111 (60.3)	107 (52.7)	218 (56.3)
Yes	73 (39.7)	96 (47.3)	169 (43.7)

[a]Baseline data are presented for the intention-to-treat sample of 387 households (184 IG and 203 CG).

[b]This is the religion of the household head.

[c]Caste categories from most to least advantaged: general, other backward caste, and scheduled caste. The scheduled tribe category is missing, as there are no tribes in the study area.

[d]HH: household.

[e]N/A: not applicable.

Table 3. Baseline characteristics of participating villages, by treatment group.

Variable[a]	Intervention	Control	All
Village population, mean (SD)[b]			
Total	974 (673.84)	1129 (1056.51)	1051 (871.79)
Total 0-6 years	166 (111.05)	188 (171.36)	177 (141.95)
Total SC[c]	225 (194.85)	422 (484.76)	323 (375.62)
Village electrification, n(%)			
No electricity	1 (7.7)	2 (15.4)	3 (11.5)
Less than 6 hours	2 (15.4)	0 (0.0)	2 (7.7)
More than 6 hours	10 (76.9)	11 (84.6)	21 (80.8)
Characteristics of participating HH[d], mean (SD)			
Number of Muslim HH per village	31 (61.39)	98 (168.96)	65 (129.25)
Number of eligible HH[e] per village	14 (7.39)	16 (12.46)	15 (10.07)
% poor[f] (Q1+Q2) per village	44.6 (0.19)	34.8 (0.23)	39.7 (0.21)
% better off (Q4+Q5) per village	39.3 (0.26)	47.5 (0.28)	43.4 (0.27)
% SC per village	31.0 (0.31)	41.5 (0.34)	36.3 (0.32)
% of mothers with 0 schooling per village	32.4 (0.17)	41.0 (0.24)	36.7 (0.21)
% of fathers with 0 schooling per village	14.6 (0.12)	9.8 (0.09)	12.2 (0.11)

[a]Baseline data are presented for the intention-to-treat sample of 26 villages (13 IG and 13 CG) containing 387 households (184 IG and 203 CG).

[b]Data from the 2011 Census of India.

[c]SC: scheduled caste (least privileged).

[d]HH: household.

[e]Eligible household: at least one child aged less than 12 months at baseline.

[f]Poor versus better off households based on asset indices (wealth quintiles).

Primary Outcomes

Ex-ante criteria were fully satisfied (Table 4). Results from 2 independent sources demonstrated a very low (1% or less) rate of contamination (Table 4; Multimedia Appendix 1). Uptake of interventions (adoption) was very high; overall participation in one or more new interventions was 94.0% (173/184), 78.3% (144/184) for the face-to-face channel, and 67.4% (124/184) for the mHealth channel (Table 4). A total of 38.0% (70/184) of participating households used the mHealth intervention weekly. Together, these results confirm the feasibility of the future main study and demonstrate the potential to impact population health.

Table 4. Primary outcomes.

Primary outcomes[a,b]	Ex-ante criteria	Ex-post results
Feasibility of the future main study		
Recruitment and randomization (villages)	70% of villages approached will agree to participate and accept randomization	100% (29/29 villages) agreed[b]
Recruitment and randomization (households)	In participating villages, 70% of eligible households will agree to participate and accept randomization	98.0% (387/395 households contacted) agreed[b]
Retention (households)	50% of households participating in the baseline survey will agree to participate in the end line survey	82.2% (318/387) enrolled households agreed[b] and 2.1% (8/387) households refused
Contamination	Contamination proportion between treatment groups should be <15%	0.6% (1/166 control end line respondents called); 0.07% (1/1310 unique callers to IVR system from a control village)[c]
Uptake (adoption)	50% of households recruited to the study will participate	94.0% (173/184) of households participated
mHealth[d] interventions	Either by listening to ≥1 mHealth item	67.4% (124/184) listened to ≥1 mHealth item
Small group meetings	Or by attending ≥1 small group meeting	78.3% (144/184) attended ≥1 meeting

[a]Feasibility outcomes were computed using the intention-to-treat (ITT) sample of 387 households (184 IG and 203 CG). Uptake was computed using the ITT intervention group sample (184 households).

[b]See flow diagram (Figure 2).

[c]See Multimedia Appendix 1.

[d]mHealth: mobile health.

Secondary Outcomes

Coverage

Uptake of the 3 intervention channels (mHealth vaccination reminders, mHealth edutainment capsules, and face-to-face small group meetings) differed among user segments (Tables 5 and 6 present modeled results; Supplementary Table 1 provides bivariate associations).

The ownership of a mobile phone was common among IG households (173/184, 94.0%) and a critical precondition for uptake of both mHealth strategies. Owing to the very few (n=11) households without a mobile phone and the prognostic importance of this variable, effect size estimates for mobile phone ownership are unreliable. However, estimates for other variables are, in principle, unbiased:

- mHealth audio vaccination reminders were accessed by 62.5% (115/184) of households. In addition to mobile phone ownership, 2 factors predicted higher uptake: high maternal education (secondary 9 years or higher vs none; OR 4.45, 95% CI 1.17-16.88; *P*=.03) and mothers' ease of access to the mobile phone (OR 3.55, 95% CI 1.08-11.71;*P*=.04).
- mHealth edutainment capsules were accessed by 60.3% (111/184) of households. In addition to mobile phone ownership, intervention uptake was predicted by high (as compared with low) women's empowerment (OR 3.29, 95% CI 1.28-8.47;*P*=.01), with some evidence of greater uptake by the lowest castes (members of scheduled castes vs general castes; OR 2.79, 95% CI 0.95-8.21;*P*=.06). However, poor phone network quality impeded the uptake of edutainment capsules (OR 0.29, 95% CI 0.12-0.71;*P*=.01).
- Face-to-face small group meetings were attended by 78.3% (144/184) of households. Living far from the meeting site reduced the uptake of small meetings (OR 0.07, 95% CI 0.02-0.33;*P* <.001); no other factor predicted uptake.

We also studied factors shaping the intensity of uptake. In modeled analyses, the number of mHealth items heard was influenced by 3 factors: mother's possession of a mobile phone, mother's ease of phone access, and women's empowerment. The number of small group meetings attended was influenced by 2 factors: living far from the meeting site and women's empowerment (Multimedia Appendix 1).

Table 5. Determinants of mobile health intervention uptake.

Variable[a,b,c,d]	Vaccination reminders		Edutainment	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Wealth quintile				
Poorest (Q1; reference)	__[e]		—	—
(Q2)	0.51 (0.12-2.15)	.36	0.42 (0.12-1.52)	.19
(Q3)	0.56 (0.21-1.51)	.26	0.78 (0.22-2.71)	.69
(Q4)	1.15 (0.36-3.64)	.83	0.74 (0.28-1.93)	.54
Highest (Q5)	0.43 (0.09-2.11)	.30	1.24 (0.40-3.92)	.71
Caste[d]				
General (reference)	—	—	—	—
Other backward caste	—	—	1.15 (0.36-3.67)	.81
Scheduled caste	—	—	2.79 (0.95-8.21)	.06
Education of mother				
None (reference)	—	—	—	—
Primary	0.80 (0.26-2.50)	.70	1.21 (0.53-2.79)	.65
Secondary or higher	4.45 (1.17-16.88)	.03	1.95 (0.56-6.80)	.29
Education of father				
None (reference)	—	—	—	—
Primary	2.01 (0.60-6.70)	.26	1.15 (0.45-2.94)	.77
Secondary or higher	2.01 (0.62-6.47)	.24	1.52 (0.52-4.44)	.45
HH[f] owns phone				
Yes versus no	23.90 (5.09-112.1)	.001	16.80 (4.27-66.18)	.001
Cell network: poor				
Yes versus no	—	—	0.29 (0.12-0.71)	.01
HH phone belonging to mother				
Yes versus no	1.21 (0.5-2.61)	.64	—	—
Mother's phone access				
Easy versus no access	3.55 (1.08-11.71)	.04	—	—
Mother can dial				
No versus yes	0.82 (0.27-2.55)	.74	—	—
Female empowerment				
Lowest (reference)	—	—	—	—
Average	—	—	0.96 (0.4-2.09)	.91
Highest	—	—	3.29 (1.28-8.47)	.01

[a]Analyses based on the intention-to-treat intervention group sample comprising 184 households.

[b]Estimates produced using Firth logistic regression with cluster bootstrapped standard errors (1000 iterations).

[c]We present the full models implemented for each outcome. Potential determinants with no evidence of association at the $P<.25$ level were not included in the models.

[d]Caste categories from most to least advantaged: general, other backward caste, and scheduled caste. The scheduled tribe category is missing, as there are no tribes in the study area.

[e]—: empty cells signify that variables were not included in models. Please see the *Methods* section on *Multivariable Analyses* subheading Coverage for further details.

[f]HH: household.

Table 6. Determinants of face-to-face intervention uptake.

Variable[a,b,c,d]	Small group meetings	
	OR (95% CI)	*P* value
Wealth quintile		
Poorest (Q1; reference)	—[e]	—
(Q2)	0.67 (0.12-3.69)	.64
(Q3)	0.60 (0.11-3.28)	.55
(Q4)	0.50 (0.08-3.00)	.45
Highest (Q5)	0.68 (0.16- 2.88)	.60
Education of mother		
None (reference)		
Primary	0.62 (0.11-3.33)	.58
Secondary or higher	0.41 (0.04- 3.87)	.44
Education of father		
None (reference)	—	—
Primary	2.76 (0.32-23.70)	.35
Secondary or higher	2.84 (0.25-31.73)	.40
HH[d] lives far		
No (reference)	—	—
Yes	0.07 (0.02-0.33)	.001
HH phone belonging to father		
Yes versus no	2.19 (0.6-8.03)	.24
HH phone belonging to someone else		
Yes versus no	0.41 (0.08-2.12)	.29
Mother can dial		
No versus yes	1.79 (0.2-15.63)	.60
Permission to attend meeting		
Yes versus no	0.69 (0.16-3.08)	.63

[a]Analyses based on the intention-to-treat intervention group sample comprising 184 households.

[b]Estimates produced using the Firth logistic regression with cluster bootstrapped standard errors (1000 iterations).

[c]We present the full models implemented for each outcome. Potential determinants with no evidence of association at the *P*=.24 level were not included in the models.

[d]HH: household.

[e]—: empty cells signify that variables were not included in models.

Adequacy of the Program Theory

Immunization knowledge was low at baseline in both study groups. For 3 of the 4 indicators studied, knowledge improved in the IG at end line (Table 7). Differences-in-differences estimates of impact suggest that observed improvements were owing to the study interventions (Table 8). Effect sizes increased after adjustment for baseline imbalances.

For 8 of 9 intermediate outcomes, the regression results showed significantly higher basic health knowledge among the IG at end line (Table 9). For one topic (whether subjects had heard of diarrhea), knowledge at end line was equal for both treatment groups (*P*=.44). This was likely owing to an independent immunization and hygiene intervention in the study area run by the Gavi Alliance and Unilever.

Table 7. Proportion of correct responses on intermediate outcomes related to immunization knowledge, by study group.

Outcome[a]	Baseline			End line		
	Treated, n (%)	Control, n (%)	*P* value[b]	Treated, n (%)	Control, n (%)	*P* value[b]
Knows immunization schedule[c]						
Yes	49 (26.6)	70 (34.5)	.095	102 (66.7)	74 (44.6)	<.001
Knows how many times to vaccinate[d]						
Correct	2 (1.1)	3 (1.5)	.734	30 (19.6)	6 (3.6)	.001
"On the vaccination card, what does each box represent?"[e]						
Correct	21 (11.4)	29 (14.3)	.400	42 (27.5)	34 (20.5)	.144
"One should vaccinate a child with a minor illness"[f]						
True	112 (60.9)	119 (58.6)	.652	110 (71.9)	95 (57.2)	.006

[a]All responses are binary.

[b]*P* value for the chi-square test of independence.

[c]This is self-assessed knowledge of the schedule from birth to 5 years.

[d]The correct response is *7 times* before age 5.

eThe correct response is a vaccine dose.

[f]The correct response is *True*.

Table 8. Impact of the intervention on intermediate outcomes related to immunization knowledge.

Outcome[a]	Model 0[b]		Model 1[c]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Knows immunization schedule from birth to 5 years	7.87 (1.90-32.49)	.004	8.40 (2.05-34.35)	.003
Knows how many times to vaccinate by age 5	3.52 (2.08-5.98)	.001	4.21 (2.25-7.85)	.001
"On the vaccination card, what does each box represent?"	1.84 (1.12-3.03)	.016	2.00 (1.18-3.40)	.011
"Children with a minor illness should be vaccinated"	1.53 (0.72-3.28)	.27	1.54 (0.71-3.34)	.27

[a]These are differences-in-differences estimates of intervention impact.

[b]Model 0=unadjusted.

[c]Model 1=adjusted for variables imbalanced at the time of randomization (wealth index and cell network).

Table 9. Estimated probability of correct responses for intermediate outcomes reflecting basic health knowledge, intervention group versus controls.

Outcomes[a]	Model 0[b]		Model 1[c]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Childhood pneumonia				
Has heard of	4.60 (2.68-7.89)	.001	4.98 (2.89-8.56)	.001
Can state signs	3.36 (1.58-7.13)	.002	3.67 (1.54-8.74)	.003
Can state how to prevent	4.12 (1.94-8.74)	.001	5.09 (2.16-12.02)	.001
Diarrhea				
Has heard of	1.24 (0.72-2.13)	.442	1.20 (0.74-1.96)	.456
Can state signs	4.14 (1.64-10.44)	.003	2.81 (1.48-5.32)	.002
Can state how to prevent	3.71 (2.06-6.67)	.001	3.82 (2.20-6.61)	.001
Dengue or chikungunya				
Has heard of	3.80 (2.35-6.10)	.001	3.97 (2.57-6.13)	.001
Can state how it is transmitted	3.61 (2.13-6.12)	.001	3.94 (2.45-6.31)	.001
Can state how to prevent	3.30 (1.97-5.53)	.001	3.53 (2.19-5.67)	.001

[a]These are basic health topics other than immunization, evaluated only at study end line.

[b]Model 0=unadjusted.

[c]Model 1=adjusted for wealth index, maternal education, paternal education, caste, and women's empowerment.

Discussion

Principal Findings

We conducted a pilot trial of SBCC interventions focusing on immunization and other basic themes important for child and family health. Interventions were delivered through small face-to-face meetings and via mobile phones using pushed audio messages and other strategies suitable for disadvantaged populations. The pilot trial aimed to assess the feasibility of a future planned main study and to draw lessons to optimize delivery of the interventions at scale.

This paper offers 4 salient findings: First, all ex-ante feasibility criteria related to recruitment, randomization, retention, and contamination were satisfied, providing compelling evidence that the planned future main trial is feasible as planned. Uptake of interventions (adoption) was near universal (50% ex-ante vs 94% in practice), demonstrating strong interest and acceptability. Second, analyses of uptake and use demonstrated that intervention use was shaped by social determinants but that the chosen combination of strategies reached all population groups, even the most vulnerable. Third, constellations of determinants differed by intervention delivery channel. Ownership of a mobile phone was critical for participation in mHealth (vaccination reminders and edutainment) interventions, whereas distance from place of residence to the meeting site was important for small group meetings. mHealth vaccination reminders were taken up preferentially by more educated women and those with easy phone access within the household, whereas mHealth edutainment capsules were favored by more empowered women and by lower caste groups, for whom the content was likely novel and useful. Face-to-face meetings were the most equitable intervention channel; participation was equal or higher among those with greater needs. Women's empowerment was an important transversal determinant, increasing uptake and intensity for all interventions. Fourth, we found that the study interventions lead to measurable improvements in basic health knowledge, supporting the potential for impact at scale. Changes in intermediate outcomes are consistent with the intervention theory of change.

Limitations

At least five important caveats should also be considered. First, an important potential bias relevant for the future definitive trial relates to vaccination coverage assessment. As documented in our pilot study and elsewhere, the population denominator used in administrative estimates is often inaccurate (and the reported number of doses unreliable) [27]. Although household surveys are considered superior, they are also affected by seasonal and chance variations and do not shed light directly on population immunity gaps. Improvements in vaccination coverage assessment methods would strengthen the ability of the main trial to assess immunization program performance. Second, those delivering and receiving interventions could not be blinded due to the nature of the interventions. We attempted to limit potential bias due to lack of blinding by hiring independent outcome assessors unaware of study purpose and group assignment, using indicators based on objective measures where possible, and triangulating between multiple measures and data sources to strengthen inference. Third, although interventions were able to achieve widespread reach in highly disadvantaged populations to ensure equity and impact, these findings are limited by the relatively small size of the IG. Fourth, quality of intervention delivery may be more difficult to achieve in a routine care setting. The personnel delivering the RCT interventions were highly motivated, well trained, and closely supervised. It may be difficult to maintain delivery quality at scale, particularly for face-to-face components. Fifth, the pilot study duration was too short (and sample size was too small)

to assess definitive changes in health outcomes. Future research is required to demonstrate whether these gains in health knowledge result in improved vaccination coverage and child health.

Generalizability

We highlight 3 insights relevant to the future definitive trial and other studies: The first relates to adapting mHealth interventions for highly disadvantaged populations. In rural India, mobile phone penetration and infrastructure is increasing rapidly, reducing barriers to delivering mHealth interventions. Our experience demonstrates that mHealth interventions can achieve reach and improve knowledge even in highly underprivileged populations, but that technical delivery and content must be substantially adapted. Although mHealth interventions using SMS have shown promise [6], we privileged audio messages with pushed dial outs owing to the low literacy, numeracy, and technological comfort levels of our target beneficiaries. We found that engaging story formats inclusive of diverse social groups were appreciated and that pure informational approaches such as vaccination reminders were taken up preferentially by the (relative) elite. As compared with SMS, audio messaging is more amenable to culture-specific contextualization and an edutainment approach. The second insight concerns how gender-related barriers shape immunization access [28] and affected the study interventions. Participation in mobile phone interventions was limited by women's ease of access to mobile phones, and, to a lesser degree, with technological familiarity. Participation in face-to-face meetings was limited by norms governing women's freedom of movement. Barriers were mitigated over time as families came to understand and value the interventions (Pérez et al, unpublished data, 2020). Unexpectedly, men, particularly fathers, were highly active participants. Future interventions should include a focus on men and families to strengthen inclusion and mitigate gender barriers. The third insight concerns the complementarity of mHealth and face-to-face communication: Although mHealth audio messaging is a promising strategy to deliver basic health information, our experience shows that it must be accompanied by face-to-face contact to enhance uptake [29] and equity. An mHealth strategy can extend the reach of face-to-face communication at high fidelity and low cost. Particularly among vulnerable populations, it is unlikely to fully replace in-person interaction. Future research exploring innovative delivery modalities while considering potential trade-offs between equity and efficiency (cost-effectiveness) is recommended.

Conclusions

A novel SBCC intervention model using face-to-face and mHealth approaches is feasible to evaluate in a future randomized trial and has the potential to strengthen the delivery of immunization and universal primary health care. The interventions achieved widespread reach in a highly disadvantaged population and showed early evidence of impact on participants' knowledge, supporting the intervention theory of change. Behavior change communication via mobile phones proved viable and contributed to standardization and scalability. Face-to-face interactions remain necessary to achieve equity and reach, suggesting the need for ongoing health system strengthening to accompany the introduction of promising mobile phone technologies.

Acknowledgments

The authors gratefully acknowledge the support of IC-IMPACTS (the India-Canada Centre for Innovative Multidisciplinary Partnerships to Accelerate Community Transformation and Sustainability), Grand Challenges Canada (Award # R-ST-POC-1707-06282), and the Canadian Institutes for Health Research (Award # 342296), who funded this study. The study sponsors played no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. The authors would like to thank Aaditeshwar Seth, Rohit Singh, Vijay Sai Pratap, Sangeeta Saini, Dinesh Rautela, and the entire Gram Vaani team (Gurgaon, India) for expert leadership of the mHealth interventions. Valuable contributions to research design and instruments were made by Claudia Bojanowski, Alexandra Otis, Myriam Cielo Perez, and Dr Isabelle Michaud-Letourneau (Université de Montréal), and Frédérique Sauvé (McGill University). Dr Bhupendra Tripathi (Bill & Melinda Gates Foundation) has been a constant source of guidance and encouragement. At the state and district levels, the authors would like to thank Mr Pankaj Kumar (Mission Director, National Health Mission, UP), Dr Arun Chaturvedi (State EPI Officer, UP), Dr Ved Prakash (General Manager, Routine Immunization, UP), Dr SK Rawat (Chief Medical Officer, Hardoi), Dr PN Chaturvedi (Former Chief Medical Officer, Hardoi), Dr Vijay Singh (District Immunization Officer, Hardoi), Dr Prakash Kumar Chaurashiya (Former District Program Officer, Hardoi), Mrs Buddhi Mishra (District Program Officer, Hardoi), Mr CK Shrivastav (Assistant Program Officer, Hardoi), Dr Prabhakar Tripathi (Medical Officer In-charge, Bawan Block, Hardoi), and Ms Gayatri Tiwari (Block Program Manager, Bawan Block, Hardoi). Study interventions were codeveloped and implemented in partnership with communities. The authors thank Mr Ajit Solanki (Chief Executive Officer of Jagriti Foundation) and the entire *Tika Vaani* field team for their dedication and commitment to community development. The authors highlight the remarkable contributions of Shailendra Singh Somvanshi, Dhiresh Singh (Chanchal), Madhavi Shukla, and Mohd. Haneef Siddiqui, Poonam Bauddh, Priyanka Mishra, and Shivam Pal.

Finally, the authors express heartfelt gratitude to the residents of Hardoi district for collaborating to improve basic health conditions for their children, families, and communities.

Authors' Contributions

MJ designed the study, acquired the funding, performed the statistical analyses and wrote the draft manuscript; DC contributed to intervention design, oversaw study implementation, contributed to data management, and revised the manuscript; GK had primary responsibility for data management, contributed to the statistical analyses, and revised the manuscript; MPS, AM, SH, and AN advised on study methodology and revised the manuscript.

All authors gave final approval of the version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods and results.
[PDF File (Adobe PDF File), 606 KB - mhealth_v8i9e20356_app1.pdf]

References

1. Sustainable Development Goals. Gavi, the Vaccine Alliance. URL: https://www.gavi.org/about/ghd/sdg/ [accessed 2020-04-01]
2. Gurnani V, Haldar P, Aggarwal MK, Das MK, Chauhan A, Murray J, et al. Improving vaccination coverage in india: lessons from intensified mission Indradhanush, a cross-sectoral systems strengthening strategy. Br Med J 2018 Dec 7;363:k4782 [FREE Full text] [doi: 10.1136/bmj.k4782] [Medline: 30530467]
3. Johri M, Rodgers L, Chandra D, Abou-Rizk C, Nash E, Mathur AK. Implementation fidelity of village health and nutrition days in Hardoi District, Uttar Pradesh, India: a cross-sectional survey. BMC Health Serv Res 2019 Oct 26;19(1):756 [FREE Full text] [doi: 10.1186/s12913-019-4625-9] [Medline: 31655588]
4. Oyo-Ita A, Wiysonge CS, Oringanje C, Nwachukwu CE, Oduwole O, Meremikwu MM. Interventions for improving coverage of childhood immunisation in low- and middle-income countries. Cochrane Database Syst Rev 2016 Jul 10;7:CD008145 [FREE Full text] [doi: 10.1002/14651858.CD008145.pub3] [Medline: 27394698]
5. Johri M, Pérez MC, Arsenault C, Sharma JK, Pai NP, Pahwa S, et al. Strategies to increase the demand for childhood vaccination in low- and middle-income countries: a systematic review and meta-analysis. Bull World Health Organ 2015 May 1;93(5):339-46C [FREE Full text] [doi: 10.2471/BLT.14.146951] [Medline: 26229205]
6. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
7. Thabane L, Cambon L, Potvin L, Pommier J, Kivits J, Minary L, Discussion Panel. Population health intervention research: what is the place for pilot studies? Trials 2019 May 30;20(1):309 [FREE Full text] [doi: 10.1186/s13063-019-3422-4] [Medline: 31146768]
8. Annual Health Survey 2010-11 Fact Sheet. Office of the Registrar General & Census Commissioner, India (ORGI). 2020. URL: http://www.censusindia.gov.in/vital_statistics/AHSBulletins/AHS_Baseline_Factsheets/U_P.pdf [accessed 2020-04-01]
9. Ram U, Jha P, Ram F, Kumar K, Awasthi S, Shet A, et al. Neonatal, 1-59 month, and under-5 mortality in 597 Indian districts, 2001 to 2012: estimates from national demographic and mortality surveys. Lancet Glob Health 2013 Oct;1(4):e219-e226 [FREE Full text] [doi: 10.1016/S2214-109X(13)70073-1] [Medline: 25104347]
10. District Level Household and Facility Survey: Uttar Pradesh. International Institute for Population Sciences (IIPS). 2010. URL: http://www.rchiips.org/pdf/rch3/report/UP.pdf [accessed 2020-04-01]
11. Pérez M, Singh R, Chandra D, Ridde V, Seth A, Johri M. Development of an mHealth Behavior Change Communication Strategy: A case-study from rural Uttar Pradesh in India. In: Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies. 2020 Presented at: COMPASS'20; June 15-17, 2020; Guayaquil, Ecuador. [doi: 10.1145/3378393.3402505]
12. Kevany S, Khumalo-Sakutukwa G, Murima O, Chingono A, Modiba P, Gray G, et al. Health diplomacy and the adaptation of global health interventions to local needs in sub-Saharan Africa and Thailand: evaluating findings from project accept (HPTN 043). BMC Public Health 2012 Jun 20;12:459 [FREE Full text] [doi: 10.1186/1471-2458-12-459] [Medline: 22716131]
13. Mobile Vaani. YouTube. URL: https://www.youtube.com/watch?v=XYsMQM50qCs [accessed 2020-04-01]
14. Mobile Vaani. Gram Vaani. URL: http://voice.gramvaani.org/vapp/mnews/912/show/tags/episode/ [accessed 2020-04-01]
15. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
16. National Department of Health. In: Peters DH, editor. Implementation Research in Health: A Practical Guide. Geneva, Switzerland: World Health Organization; 2013.

17. Kaufman J, Ames H, Bosch-Capblanch X, Cartier Y, Cliff J, Glenton C, et al. The comprehensive 'communicate to vaccinate' taxonomy of communication interventions for childhood vaccination in routine and campaign contexts. BMC Public Health 2017 May 10;17(1):423 [FREE Full text] [doi: 10.1186/s12889-017-4320-x] [Medline: 28486956]
18. DHS Model Questionnaire. The DHS Program. 2008. URL: http://dhsprogram.com/publications/publication-dhsq6-dhs-questionnaires-and-manuals.cfm [accessed 2020-04-25]
19. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. Demography 2001 Feb;38(1):115-132. [doi: 10.1353/dem.2001.0003] [Medline: 11227840]
20. Sample Size Calculator. URL: http://www.sample-size.net/ [accessed 2020-04-25]
21. CD Block Wise Primary Census Abstract Data (PCA) - Uttar Pradesh. Office of the Registrar General & Census Commissioner, India (ORGI). 2014. URL: http://censusindia.gov.in/pca/cdb_pca_census/Houselisting-housing-UP.html [accessed 2020-04-25]
22. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2008. URL: https://www.r-project.org/about.html [accessed 2020-03-01]
23. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Making 2001;21(1):45-56. [doi: 10.1177/0272989X0102100106] [Medline: 11206946]
24. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 2000 Apr 30;19(8):1059-1079. [doi: 10.1002/(sici)1097-0258(20000430)19:8<1059::aid-sim412>3.0.co;2-0] [Medline: 10790680]
25. OMODEL: Stata Modules to Perform Tests on Ordered Probit and Ordered Logit Models. EconPapers. 1997. URL: https://EconPapers.repec.org/RePEc:boc:bocode:s320901 [accessed 2020-04-01]
26. Khandker S, Koolwal G, Hussain AS. Handbook on Impact Evaluation: Quantitative Methods and Practices. Geneva, Switzerland: The World Bank; 2010.
27. Cutts FT, Claquin P, Danovaro-Holliday MC, Rhoda DA. Monitoring vaccination coverage: defining the role of surveys. Vaccine 2016 Jul 29;34(35):4103-4109 [FREE Full text] [doi: 10.1016/j.vaccine.2016.06.053] [Medline: 27349841]
28. Feletto M, Sharkey A. The influence of gender on immunisation: using an ecological framework to examine intersecting inequities and pathways to change. BMJ Glob Health 2019;4(5):e001711 [FREE Full text] [doi: 10.1136/bmjgh-2019-001711] [Medline: 31565415]
29. Chakraborty D, Gupta A. Seth Axperiences From a Mobile-based Behaviour Change Campaign on Maternal and Child Nutrition in Rural India. In: Proceedings of the Tenth International Conference on Information and Communication Technologies and Development. 2019 Presented at: ICTD'19; January 4-7, 2019; Ahmadabad, India URL: https://doi.org/10.1145/3287098.3287110 [doi: 10.1145/3287098.3287110]

Abbreviations

ANM: Assistant Nurse Midwives
ASHA: Accredited Social Health Activists
AWW: Anganwadi Workers
CG: control group
CTRI: Clinical Trials Registry–India
GoI: Government of India
GEE: generalized estimating equations
IG: intervention group
ITT: intention-to-treat
IVR: interactive voice response
LMIC: low- and middle-income country
mHealth: mobile health
RCT: randomized controlled trial
SBCC: social and behavior change communication
SDG: Sustainable Development Goals
UN: United Nations
UP: Uttar Pradesh
VHND: Village Health and Nutrition Day
WASH: water, sanitation, and hygiene

Edited by G Eysenbach; submitted 21.05.20; peer-reviewed by S Kevany, F Palmieri; comments to author 27.06.20; revised version received 01.07.20; accepted 22.07.20; published 21.09.20.

Please cite as:
Johri M, Chandra D, Kone KG, Sylvestre MP, Mathur AK, Harper S, Nandi A
Social and Behavior Change Communication Interventions Delivered Face-to-Face and by a Mobile Phone to Strengthen Vaccination Uptake and Improve Child Health in Rural India: Randomized Pilot Study
JMIR Mhealth Uhealth 2020;8(9):e20356
URL: http://mhealth.jmir.org/2020/9/e20356/
doi:10.2196/20356
PMID:32955455

Original Paper

SMS Text Messages for Parents for the Prevention of Child Drowning in Bangladesh: Acceptability Study

Md Mosharaf Hossain[1*], PhD; Kulanthayan Mani[2*], PhD; Ruhani Mat Min[1], PhD

[1]Faculty of Business, Economics and Social Development, Universiti Malaysia Terengganu, Kuala Nerus, Terengganu, Malaysia

[2]Safe Kids Malaysia and Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia

[*]these authors contributed equally

Corresponding Author:
Kulanthayan Mani, PhD
Safe Kids Malaysia and Department of Community Health
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
43400 Serdang
Selangor
Malaysia
Phone: 60 0397692398
Email: kulan@upm.edu.my

Abstract

Background: In many cases, greater use is being made of mobile phone text messages as a means of communication between patients and health care providers in countries around the world.

Objective: We studied the use of mobile phones and the factors related to the acceptability of text messages for parents for the prevention of child drowning in Bangladesh.

Methods: From a randomized controlled trial involving 800 parents, 10% (80/800) were selected, and socioeconomic status, mobile phone use, and acceptability of SMS text messages for drowning prevention were measured. Participants with at least one child under 5 years of age were selected from rural areas in Rajshahi District in Bangladesh. Mobile phone–based SMS text messages were sent to the participants. Multivariate regression was used to determine the factors related to the acceptability of text messages for the prevention of child drowning in Bangladesh.

Results: The acceptability of SMS text messages for the prevention of child drowning in Bangladesh was significantly lower among women (odds ratio [OR] 0.50, 95% CI 0.12-1.96, *P*=.02) than among men, lower for parents older than 30 years (OR 0.17, 95% CI 0.14-1.70, *P*=.01) compared to parents younger than 30 years, higher among parents who had an education (OR 1.63, 95% CI 1.11-5.80, *P*=.04) than among illiterate parents, and higher among parents with a monthly household income over 7000 Bangladeshi Taka (approximately US $82.54; OR 1.27, 95% CI 1.06-1.96, *P*=.05) than among parents whose monthly income was less than 7000 Bangladeshi Taka.

Conclusions: The high percentage of mobile phone use and the acceptability of SMS text messages for parents for the prevention of child drowning are encouraging, in terms of identifying the best strategy for using such technologies, and deserve further evaluation.

(JMIR Mhealth Uhealth 2020;8(9):e16958) doi:10.2196/16958

KEYWORDS

acceptability; SMS; drowning; parents

Introduction

Water plays an important role in children's daily life. They like to play with water, have fun, and sometimes are more adventurous. Children are always excited around water, no matter whether it is a pond, pool, lake, in an open field, or simply on or beside the road after rain falls. It is impossible for children to grow up without water, as besides playing, they need it to clean themselves, and they gain comfort and are cooled by it. Although water is considered to be an important element for children to survive, it is also a hazard if there is a lack of awareness concerning the dangers it presents [1].

A small child can drown in only a few inches of water—under a bucket, in a field, or in the bathtub. Drowning injuries have an epidemiological pattern. However, the pattern changes depending on the type of water body, age group, and activity. Drowning ranks among the top 3 causes of accidental deaths in most countries around the world, and the mortality rates are highest in children under 5 years of age [1].

Drowning is the leading cause of death for children aged 1 year and over in Bangladesh. According to the rate reported in the Bangladesh Health and Injury Survey [2], nearly 17,000 children drowned in 2004, approximately 46 per day, and close to 4 times more than the rate in other low-income countries [2]. This equates to approximately 188 children drowning a day. In Bangladesh, drowning rates are 10 to 20 times higher than those in high-income countries [2]. There is a question concerning whether drowning is always fatal or whether it is possible to survive. While many people think that drowning is always a fatal event, others think that near drowning is a result that is never lethal [2].

As a result, drowning injuries are now the leading cause of death, disability, and severe morbidity among children in low- and middle-income countries in Asia, such as Bangladesh. This compromises the gains that were previously achieved, at high cost, to prevent other causes of illness and injury among children, and it jeopardizes the continued progress in survival and security. One conclusion is that drowning injuries are so important that they must be treated first [3].

Drowning reduces the impact of other interventions for children, as children who died from drowning often received conventional vaccines, vitamin A supplementation, and other food aid. They often used early development and education programs, and such investments are lost when a child drowns. Effective drowning prevention interventions are available at all stages of childhood and are cost-effective alongside traditional interventions for children [2].

Mobile phones and SMS text messages are gradually being used between patients and their health care providers in many countries around the world because of their communication potential [4]. Several studies [5-7] have shown that SMS text message–based counseling and observation can restore patient behavior and health outcomes. However, the potential use of SMS text messages in clinical settings in low-income countries has not been well established. The feasibility of implementing such technologies is not clearly demonstrated, nor is there a cost-effective and sustainable payment mechanism or commercial model that can be improved [8-12]. Information concerning how patients appreciate or perceive the use of SMS text messages from a mobile phone to improve disease management can be valuable in implementing such methods in primary care. Therefore, this study aimed to determine the access to mobile phones and the willingness to receive SMS text messages on a mobile phone; and determine the factors related to the willingness to receive mobile phone–based SMS text messages concerning child drowning prevention.

Methods

Overview

The researchers focused on the acceptability of text messages for child drowning prevention as part of a randomized controlled trial (International Standard Randomized Controlled Trial Number; ISRCTN13774693) of a mobile phone–based SMS text messages intervention aimed at improving the knowledge, attitude, and practices of parents about child drowning in Bangladesh. As part of a larger study, the design of the study and the data collection procedure followed the protocol of a previously published randomized controlled trial [13]. The sample size was 80 parents from 2 villages (10% of the randomized controlled trial's sample size). The parents received an SMS text message before the interview.

The mobile phone–based SMS text message intervention aimed to improve parents' knowledge, attitudes, and practices in relation to the prevention of child drowning in Bangladesh. The researcher developed the SMS text messages based on focus group discussions and literature reviews. The messages used informal language and were sent every Friday. On Friday morning, each week, the research team sent out SMS text messages to parents in the intervention group. The message was typically 150-200 characters long and in the Bangla language.

Data were collected through face-to-face interviews in rural areas using a structured questionnaire, which was conducted by a team of 2 qualified research assistants. The inclusion criteria were the same as those for the published study protocol [13]. Parents meeting the selection criteria were included in the study. The questionnaires included questions on socioeconomic characteristics (age, sex, marital status, education, and income), mobile phone use, SMS text message use, and drowning knowledge.

Data Analysis

The researchers performed chi-square tests to assess the importance of the responses and associated factors regarding SMS text messaging concerning drowning in childhood. A univariate logistic regression analysis was used to evaluate the relationship between SMS text message reading and the individual variables. Factors having a statistically significant association with the reading of text messages on child drowning in univariate analysis (P value<.25) were included in the final logistic model (P value<.05) taking into account other variables. SPSS statistical software (version 22; IBM Corp) was used for data analysis. A value of P<.05 was considered statistically significant.

Ethical Statement

The ethical approval for this study was obtained from University Putra Malaysia (UPM/TNCPI/RMC/1.4.18.1 (JKEUPM)/F2) and the Centre for Injury Prevention and Research, Bangladesh. Informed written agreements were obtained from all the respondents before the data were collected. All the respondents were assured that the data would only be used for research purposes and that all the answers would be kept confidential.

Results

The association between the acceptability of text messages about childhood drowning and the selected sociodemographic characteristics of parents are presented in Table 1. Table 1 reveals the associations between parent age, gender, educational status, occupation, monthly income, the ability to read SMS text messages, liking SMS text messages, and having Bangla language on their phone with the acceptability of text messages on childhood drowning. The acceptability of text messages concerning childhood drowning prevention was significantly higher among males (76%, χ^2=4.39, *P*<.001), parents aged less than 30 years (82%, χ^2=6.54, *P*<.001), literate parents (76%, χ^2=4.98, *P*<.001), parents with a monthly income of more than 7000 BDT (80%, χ^2=5.69, *P*<.001), parents who had the ability to read an SMS text messages (71%, χ^2=3.94, *P*<.001), parents who liked mobile phone–based SMS text messages (74%, χ^2=7.51, *P*<.001), and parents who had Bangla on the mobile phone (73%, χ^2=5.63, *P*<.001) (Table 1). There were no significant associations between parents' occupation, type of mobile phone, use of mobile phone, knowledge about drowning with the acceptability of text messages on childhood drowning.

Table 1. The results of chi-square tests concerning the acceptability of text messages about childhood drowning among various sociodemographic variables in Bangladesh.

Variables	All (N=80), n (%)	Acceptability		Comparison	
		No (n=26), n (%)	Yes (n=54), n (%)	Chi-square (*df*)	*P* value
Age (years)				6.54 (1,80)	<.001
Below 30	38 (47)	7 (18)	31 (82)		
More than 30	42 (53)	19 (45)	23 (55)		
Gender				4.39 (1,80)	<.001
Male	50 (63)	12 (24)	38 (76)		
Female	30 (37)	14 (47)	16 (53)		
Educational status				4.98 (1,80)	<.001
Illiterate	30 (37)	14 (47)	16 (53)		
Literate	50 (63)	12 (24)	38 (76)		
Occupation				0.97 (1,80)	.563
Housewife/Farmer	66 (83)	21 (38)	38 (62)		
Other	14 (17)	5 (17)	16 (83)		
Monthly income (BDT[a])				5.69 (1,80)	.001
Less than 7000	40 (50)	8 (20)	22 (55)		
More than 7000	40 (50)	18 (45)	32 (80)		
Phone type				0.38 (1,80)	.781
Normal	68 (85)	11 (37)	19 (63)		
Smartphone	12 (15)	15 (30)	35 (70)		
Phone use				0.36 (1,80)	.144
Father	30 (37)	21 (38)	45 (66)		
Mother	50 (63)	5 (29)	9 (71)		
Ability to read SMS text messages				3.94 (1,80)	<.001
No	10 (12)	6 (50)	4 (50)		
Yes	70 (88)	20 (29)	50 (71)		
Liked SMS text messages				7.51 (1,80)	<.001
No	12 (15)	8 (66)	4 (34)		
Yes	68 (85)	18 (26)	50 (74)		
Have Bangla language on phone				5.63 (1,80)	<.001
No	11 (14)	7 (63)	4 (37)		
Yes	69 (86)	19 (27)	50 (73)		
Knowledge about drowning				3.94 (1,80)	.346
No	10 (12)	6 (50)	4 (50)		
Yes	70 (88)	20 (29)	50 (71)		

[a]BDT: Bangladeshi Taka; an exchange rate of US $1 to 0.012 BDT is applicable.

In the logistic analysis (Table 2), the acceptability of text messages on childhood drowning was significantly higher among males. The odds ratio (OR) of 0.50 for females was less than 1, indicating that female respondents considered text messages on childhood drowning to be 0.50 (adjusted OR 0.50, *P*=.02, 95% CI 0.12-1.96) times less acceptable than their male counterparts.Parents aged more than 30 years considered text messages on childhood drowning to be 0.17 times less acceptable than parents aged less than 30 years (OR 0.17, *P*=.01, 95% CI 0.14-1.70). Parents who had an education considered text messages on childhood drowning to be 1.63 times (OR 1.63, *P*=.04, 95% CI 1.11-5.80) less acceptable than parents who had no education, those with a household income of more than 7000 BDT considered text messages on childhood drowning to be 1.27 times more acceptable than those with a household income of less than 7000 BDT (OR 1.27, *P*=.05,

95% CI 1.06-1.96), and parents who had Bangla on their mobile phone considered text messages on childhood drowning to be 6.46 times more acceptable than parents who did not have the Bangla language on their mobile phone (OR 6.46, *P*=.05, 95% CI 1.89-6.65) (Figure 1). There were no significant associations among parents' occupation (OR 7.92, *P*=.052, 95% CI 0.97-64.33) or knowledge about drowning (OR 3.73, *P*=.05, 95% CI 0.95-14.71) with the acceptability of text messages on childhood drowning.

Table 2. Results of the logistic analysis of the acceptability of text messages on childhood drowning among various sociodemographic variables.

Variable[a]	OR[b]	*P* value	95% CI	Adjusted OR	*P* value	95% CI
Age (years)						
Below 30[c]	1	—	—	1	—	—
More than 30	0.27	<.001	0.09-0.75	0.17	<.001	0.14-1.70
Gender						
Male[c]	1	—	—	1	—	—
Female	0.36	.06	0.13-0.94	0.5	.02	0.12-1.96
Educational status						
Illiterate[c]	1	—	—	1	—	—
Literate	2.77	.03	1.05-7.29	1.63	.04	1.11-5.80
Occupation						
Housewife/Farmer[c]	1	—	—	1	—	—
Other	7.92	0.052	0.97-64.33	1.73	.16	1.71-8.87
Monthly income (BDT[d])						
Less than 7000[c]	1	—	—	1	—	—
More than 7000	1.30	<.001	0.98-1.82	1.27	.06	1.06-1.96
Ability to read SMS text messages						
No[c]	1	—	—	1	—	—
Yes	3.75	.05	0.95-14.71	1.04	.98	0.08-12.91
Liked SMS text messages						
No[c]	1	—	—	1	—	—
Yes	5.55	<.001	1.49-20.70	1.38	.78	0.13-13.87
Have Bangla language on phone						
No[c]	1	—	—	1	—	—
Yes	4.6	.02	1.20-17.53	6.46	.053	1.89-6.65
Knowledge about drowning						
No[c]	1	—	—	1	—	—
Yes	3.73	.05	0.95-14.71	1.26	.99	0.10-9.30

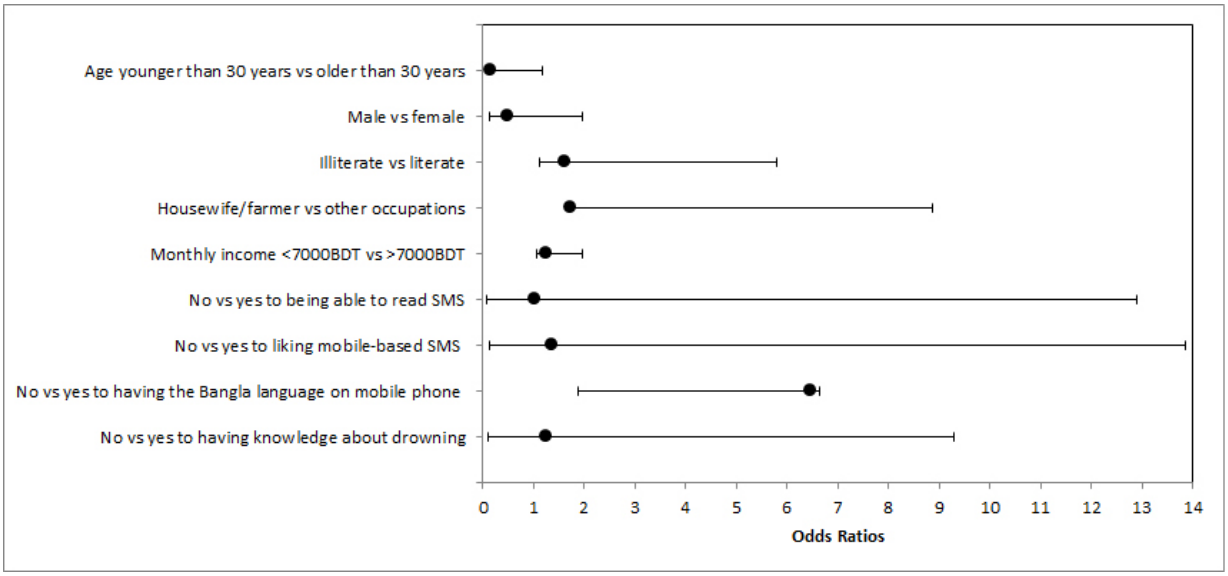
[a] Model Summary—2-loglikelihood: 71.12; Cox & Snell R Square: 0.31; Negelikereke R: 0.43; model chi-square: 29.77.

[b]OR: odds ratio.

[c]Reference.

[d]BDT: Bangladeshi Taka; an exchange rate of US $1 to 0.012 BDT is applicable.

Figure 1. Forest plot of acceptability of text messages for parents on childhood drowning prevention odds ratios.



Discussion

Principal Findings

The aim of this research was to identify access to mobile phones, the willingness to receive mobile phone–based SMS text messages in relation to the prevention of childhood drowning, and associated factors among parents of children under 5 in the rural areas of Bangladesh. The results of this study show that there is high (54/80, 68%) mobile phone access among parents with children less than 5 years of age. Of the respondents who had access to mobile phones, the majority (59/80, 74%) would like to receive phone-based SMS text messages. This is similar to the findings of a study conducted in northwest Ethiopia concerning mobile health services among patients with diabetes by Jemere et al [14] in which 71% were willing to receive mobile phone–based health services.

Age, gender, educational status, monthly income, ability to read SMS text messages, liking mobile phone–based SMS text messages, and having Bangla language software on their mobile phone were associated with the acceptability of text messages on childhood drowning prevention in Bangladesh. The majority (58/80, 72%) were able to read SMS text messages. In the study, those in the age group of less than 30 years showed a higher acceptability of receiving text messages on childhood drowning than those in the other age group (≥30 years); the acceptability of text messages on childhood drowning decreased slightly with age. Older respondents were less likely to possess the technology and knowledge compared to their younger counterparts, which could be the reason for the lower acceptability of text messages concerning childhood drowning. This is similar to the findings of a study in Bangladesh conducted by Islam et al [15] in which all the participants owned a mobile phone and about half of the participants reported being able to retrieve and read SMS text messages.

Among males, only those with a household income above 7000 BDT, and parents who had the Bangla language on their mobile phone were associated with the acceptability of text messages on childhood drowning prevention. An increase in education is associated with an increase in the proportion of income earners per household and had an impact on the acceptability of text messages on childhood drowning prevention, which is contrary to the assertions in the literature in this regard [15]. Those with a relatively higher income had greater willingness to read text messages on childhood drowning prevention, which is aligned with the expectations [15]. The association is likely to be due to the fact that a higher educational status is likely to lead to improved knowledge and awareness of drowning, and better access to mobile phone and mobile phone networks. Awareness concerning how to apply these new technologies to consumer health and enable patients to take control and play an active role in managing their health has increased. Consumer health interventions have been used, for example, to help people monitor their own health [16], provide social information and support, and remotely monitor their home [17]. Many studies [18-20] have used mobile phone technology for public health issues and have commented on other considerations for older people in terms of design, usability and functionality. Studies [18] have shown that interventions need to take into account psychological assumptions or barriers for use among older people.

Kotani et al [19] noted the success of using mobile phones to take pictures, even when older people refused to take pictures with traditional cameras. Mobile phones and telephones were perceived as informal and ubiquitous and were therefore considered acceptable while traditional cameras were not [20].

Mobile phone interventions for the health of the elderly are in their infancy and are just beginning to develop. The rapid development of mobile phones, coupled with the rapid aging of the population, provides an excellent opportunity to use mobile phone technology to better manage the health of seniors and positively impact their quality of life and well-being. Early childhood research in the field allow interested researchers to conduct research in all directions while pursuing the same goal of improving the lives of older people [21,22]. The prevalence of mobile phones, text messages, and in particular, the growing use of instant messengers and social networking apps in health care highlight ways to significantly integrate health care into

users' daily activities instead of artificially complementing health care processes [23].

There are clear benefits from these interventions, such as further educating parents about their treatment, providing more opportunities for health care professionals and parents to interact efficiently and effectively, and improving health literacy in general. Directed and monitored interventions also provide end users with reliable information in a condensed and easily accessible form. More research is required to optimize the use of these interventions, with attention particularly focused on sensitive areas, such as cost and time burden for the providers and consumers of mobile phone health care services. The study participants in this trial mostly had (more than 87%) access to a mobile phone, which might be different from other populations in Bangladesh [24-27].

Strengths and Limitations

This study provides quality evidence and establishes an association between the intervention and outcomes of the study in Bangladesh among parents of children aged under 5 years old.

These findings are consistent with the results reported by several researchers in similar studies [13-15] around the world concerning different injuries and provide information that is critical for controlling the drowning epidemic, especially in low- and middle-income countries that are highly burdened by such events.

The mobile phone intervention used in this study appears to be a relatively cheap and acceptable strategy for improving the drowning prevention knowledge, attitude, and practices of parents of children under 5 years old. The data concerning the cost-effectiveness of this strategy in comparison to those of other interventions were not reported because it was beyond the scope of this study.

Conclusion

The results of this study show that the vast majority of parents in rural areas in Bangladesh find SMS text messages to prevent drowning acceptable. Given that the number of drowning deaths is still high and the low cost of mobile phones in Bangladesh, a self-sustaining business model in low-income countries and other middle-income countries is possible. The age of the respondents, gender, educational status, monthly income, ability to read SMS text messages, liking mobile phone–based SMS text messages, and Bangla language software on the mobile phone were associated with the acceptance of text messages for preventing drowning. Based on this outcome, mobile phone SMS text messages implementation, such as self-monitoring, behavioral counseling, and interventions can be used to improve knowledge, attitudes, and practices.

Authors' Contributions

KM, MMH, and RMM wrote and approved the final manuscript. KM and MMH conceptualized and designed the research. KM and MMH undertook the formal data analysis and are responsible for all analysis. All authors critically revised this manuscript.

Conflicts of Interest

None declared.

References

1. Peden M. , Oyegbite K. Ozanne-Smith J World report on child injury prevention. URL: https://apps.who.int/iris/bitstream/handle/10665/43851/9789241563574_eng.pdf;jsessionid=664C97FA1988A066A851587670D05633?sequence=1
2. Bangladesh Health and Injury Survey. Bangladesh Health and Injury Survey. Dhaka, Bangladesh: Government of the People's Republic of Bangladesh; 2005.
3. Rahman A. Bangladesh Health and Injury Survey: Report on Children. Dhaka, Bangladesh: Government of the People's Republic of Bangladesh, ICMH, UNICEF and TASC; 2005.
4. Chai SJ, Tan F, Ji Y, Wei X, Li R, Frost M. Community-Level Text Messaging for 2009 H1N1 Prevention in China. American Journal of Preventive Medicine 2013 Aug;45(2):190-196. [doi: 10.1016/j.amepre.2013.03.014]
5. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
6. Burner ER, Menchine MD, Kubicek K, Robles M, Arora S. Perceptions of successful cues to action and opportunities to augment behavioral triggers in diabetes self-management: qualitative analysis of a mobile intervention for low-income Latinos with diabetes. J Med Internet Res 2014;16(1):e25 [FREE Full text] [doi: 10.2196/jmir.2881] [Medline: 24476784]
7. Nhavoto JA, Grönlund A. Mobile technologies and geographic information systems to improve health care systems: a literature review. JMIR Mhealth Uhealth 2014;2(2):e21 [FREE Full text] [doi: 10.2196/mhealth.3216] [Medline: 25099368]
8. Brown S, Saint M. Value proposition for mhealth monitoring solution of diabetes. 2013 Presented at: 2013 IST-Africa Conference & Exhibition; May 29-31, 2013; Nairobi, Kenya p. 1-11 URL: https://www.researchgate.net/publication/287396347_Value_proposition_for_mhealth_monitoring_solution_of_diabetes
9. Okuboyejo S, Eyesan O. mHealth: Using Mobile Technology to Support Healthcare. OJPHI 2014 Feb 05;5(3). [doi: 10.5210/ojphi.v5i3.4865]

10. Bhuiyan T, Mahmud I, Alam B. Usability analysis of sms alert system for immunization in the context of Bangladesh. IJRET 2013 Nov 25;02(11):300-305. [doi: 10.15623/ijret.2013.0211046]
11. Abebe N. Txt4Health: using mobile technology in public health communication and education campaigns. 2013 Presented at: National Conference on Health Communication, Marketing, and Media; 2013; Atlanta, GA.
12. Castaño PM, Stockwell MS, Malbon KM. Using digital technologies to improve treatment adherence. Clin Obstet Gynecol 2013 Sep;56(3):434-445. [doi: 10.1097/GRF.0b013e3182988a3b] [Medline: 23722918]
13. Hossain M, Mani KKC, Sidik SM, KS H, Fazlur Rahman A. Randomized controlled trial on drowning prevention for parents with children aged below five years in Bangladesh: a study protocol. BMC Public Health 2015 May 11;15(1). [doi: 10.1186/s12889-015-1823-1]
14. Jemere AT, Yeneneh YE, Tilahun B, Fritz F, Alemu S, Kebede M. Access to mobile phone and willingness to receive mHealth services among patients with diabetes in Northwest Ethiopia: a cross-sectional study. BMJ Open 2019 Jan 24;9(1):e021766. [doi: 10.1136/bmjopen-2018-021766]
15. Shariful ISM, Lechner A, Ferrari U, Seissler J, Holle R, Niessen LW. Mobile phone use and willingness to pay for SMS for diabetes in Bangladesh. J Public Health (Oxf) 2016 Mar;38(1):163-169. [doi: 10.1093/pubmed/fdv009] [Medline: 25687131]
16. Lam LT, Yang L, Liu Y, Geng WK, Liu DC. Different quantitative measures of the impact of injury deaths on the community in the Guangxi Province, China. Accid Anal Prev 2005 Jul;37(4):761-766. [doi: 10.1016/j.aap.2005.03.016] [Medline: 15949464]
17. Hossain M, Mani KKC, Sidik SM, Hayati KS, Rahman AKMF. Socio-demographic, environmental and caring risk factors for childhood drowning deaths in Bangladesh. BMC Pediatr 2015 Sep 10;15:114 [FREE Full text] [doi: 10.1186/s12887-015-0431-7] [Medline: 26357879]
18. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008;8:45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]
19. Kotani K, Morii M, Asai Y, Sakane N. Application of mobile-phone cameras to home health care and welfare in the elderly: experience in a rural practice. Aust J Rural Health 2005 Jun;13(3):193-194. [doi: 10.1111/j.1440-1854.2005.00682.x] [Medline: 15932491]
20. Faucounau V, Riguet M, Orvoen G, Lacombe A, Rialle V, Extra J, et al. Electronic tracking system and wandering in Alzheimer's disease: a case study. Ann Phys Rehabil Med 2009;52(7-8):579-587 [FREE Full text] [doi: 10.1016/j.rehab.2009.07.034] [Medline: 19744906]
21. Joe J, Demiris G. Older adults and mobile phones for health: a review. J Biomed Inform 2013 Oct;46(5):947-954 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.008] [Medline: 23810858]
22. Hossain MM, Mani KKC, Islam MR. Prevalence and determinants of the gender differentials risk factors of child deaths in Bangladesh: evidence from the Bangladesh demographic and health survey, 2011. PLoS Negl Trop Dis 2015 Mar;9(3):e0003616 [FREE Full text] [doi: 10.1371/journal.pntd.0003616] [Medline: 25747178]
23. Kamel E, Emam, SASameh E. Knowledge, attitude and practice among rural mothers about Home-related injuries in a rural area in El-Minia Governorate, Egypt. Science Journal of Public Health 2014;2(6):653-659. [doi: 10.11648/j.sjph.20140206.35]
24. Islam MZ. 87pc households use mobile. The Daily Star News. 2016 Jan 12. URL: https://www.thedailystar.net/frontpage/87pc-households-use-mobile-200656
25. Tang S, Ghose B, Hoque MR, Hao G, Yaya S. Women Using Mobile Phones for Health Communication Are More Likely to Use Prenatal and Postnatal Services in Bangladesh: Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e10645 [FREE Full text] [doi: 10.2196/10645] [Medline: 30816850]
26. Krinsky NI. Membrane antioxidants. Ann N Y Acad Sci 1988;551:17-32; discussion 32. [doi: 10.1111/j.1749-6632.1988.tb22317.x] [Medline: 3245661]
27. Hossain M, Mani KKC, Mohd Sidik S, Kadir Shahar H. The development of an intervention package to prevent children under five years old drowning in rural Bangladesh. Acta Paediatr 2016 Aug;105(8):e373-e378. [doi: 10.1111/apa.13447] [Medline: 27111395]

Abbreviations

OR: odds ratio

Edited by G Eysenbach; submitted 07.11.19; peer-reviewed by D Ramesh, MA(Barry; comments to author 23.01.20; revised version received 24.02.20; accepted 23.03.20; published 23.09.20.

Please cite as:
Hossain MM, Mani K, Mat Min R
SMS Text Messages for Parents for the Prevention of Child Drowning in Bangladesh: Acceptability Study
JMIR Mhealth Uhealth 2020;8(9):e16958
URL: http://mhealth.jmir.org/2020/9/e16958/
doi:10.2196/16958
PMID:32965224

Original Paper

Potential Benefits and Risks Resulting From the Introduction of Health Apps and Wearables Into the German Statutory Health Care System: Scoping Review

Alexandra Heidel[1]; Christian Hagist[1]

Chair of Economic and Social Policy, WHU – Otto Beisheim School of Management, Vallendar, Germany

Corresponding Author:
Alexandra Heidel
Chair of Economic and Social Policy
WHU – Otto Beisheim School of Management
Burgplatz 2
Vallendar, 56179
Germany
Phone: 49 261 6509 256
Email: alexandra.heidel@whu.edu

Abstract

Background: Germany is the first country worldwide that has introduced a digital care act as an incentive system to enhance the use of digital health devices, namely health apps and wearables, among its population. The act allows physicians to prescribe statutory financed and previously certified health apps and wearables to patients. This initiative has the potential to improve treatment quality through better disease management and monitoring.

Objective: The aim of this paper was to outline the key concepts related to the potential risks and benefits discussed in the current literature about health apps and wearables. Furthermore, this study aimed to answer the research question: Which risks and benefits may result from the implementation of the digital care act in Germany?

Methods: We conducted the scoping study by searching the databases PubMed, Google Scholar, and JMIR using the keywords health apps and wearables. We discussed 55 of 136 identified articles published in the English language from 2015 to March 2019 in this paper using a qualitative thematic analysis approach.

Results: We identified four key themes within the articles: Effectivity of health apps and wearables to improve health; users of health apps and wearables; the potential of bring-your-own, self-tracked data; and concerns and data privacy risks. Within these themes, we identified three main stages of benefits for the German health care system: Usage of health apps and wearables; continuing to use health apps and wearables; and sharing bring-your-own; self-tracked data with different agents in the health care sector.

Conclusions: The digital care act could lead to an improvement in treatment quality through better patient monitoring, disease management, personalized therapy, and better health education. However, physicians should play an active role in recommending and supervising health app use to reach digital-illiterate or health-illiterate people. Age must not be an exclusion criterion. Yet, concerns about data privacy and security are very strong in Germany. Transparency about data processing should be provided at all times for continuing success of the digital care act in Germany.

(JMIR Mhealth Uhealth 2020;8(9):e16444) doi:10.2196/16444

KEYWORDS

health apps; wearables; digital health application; mHealth

Introduction

Health apps and wearables have experienced increasing popularity in recent years [1]. Health apps and wearables are able to contribute more to the health care system than monitoring physical exercise, heart rate, or calories; they may support chronically ill patients with the management of specific diseases such as Parkinson's disease, diabetes, tinnitus, or stress-related symptoms. Yet, Kotlikoff and Hagist [2] outlined already in 2009 that constantly increasing health care expenditure is one of the major social challenges for modern welfare states. Health apps and wearables might hold significant potential to decrease

these costs. Apps and wearables are considered beneficial in the fields of preventive medicine and disease monitoring because the gamification of health enhances personal motivation and coordination. Germany has just launched one of the most progressive pilot projects in its health care history. The parliament passed the Digitale Versorgung Gesetz (DVG; digital care act) in 2019, which introduces the digitale Gesundheitsanwendungen (DIGA; digital health applications) into the German statutory health care system [3]. The DVG enables physicians to prescribe health apps for smartphones or wearables, which are covered for the insured by the sickness funds. This incentive system to introduce mobile health (mHealth) into the health care system is unique and exceptional worldwide [4]. The German Ministry of Health has shaped a completely new concept with the term DIGA. DIGA is a medical device within the scope of the European medical device regulation and classified as risk level I and not higher than a risk level IIa [5]. DIGA is a portable technology with the medical scope of monitoring, treatment, or reducing the effects of diseases [5]. Simple nutrition or menstrual cycle apps without any clear scope to improve the treatment effectivity of a medical condition are, for now, not considered as DIGAs.

Researchers in Germany are currently discussing the potential success of the act and the expected patient demand and acceptance. Experience with a regulation such as the DVG does not exist. According to a study by GfK, about 28% of Germans (25% female, 30% male) track at least one health parameter [6], and the average use from all 16 surveyed countries is 33%. Reasons to not track personal health data might be related to data security concerns, the accessibility of technology, or personal attitudes towards the recording of fitness parameters. We aimed to identify key concepts of the inclusion of health apps and wearables in the German statutory health care sector. We analyzed 55 of 136 identified articles to answer the research question: Which risks and benefits may result from the implementation of the digital care act in Germany?

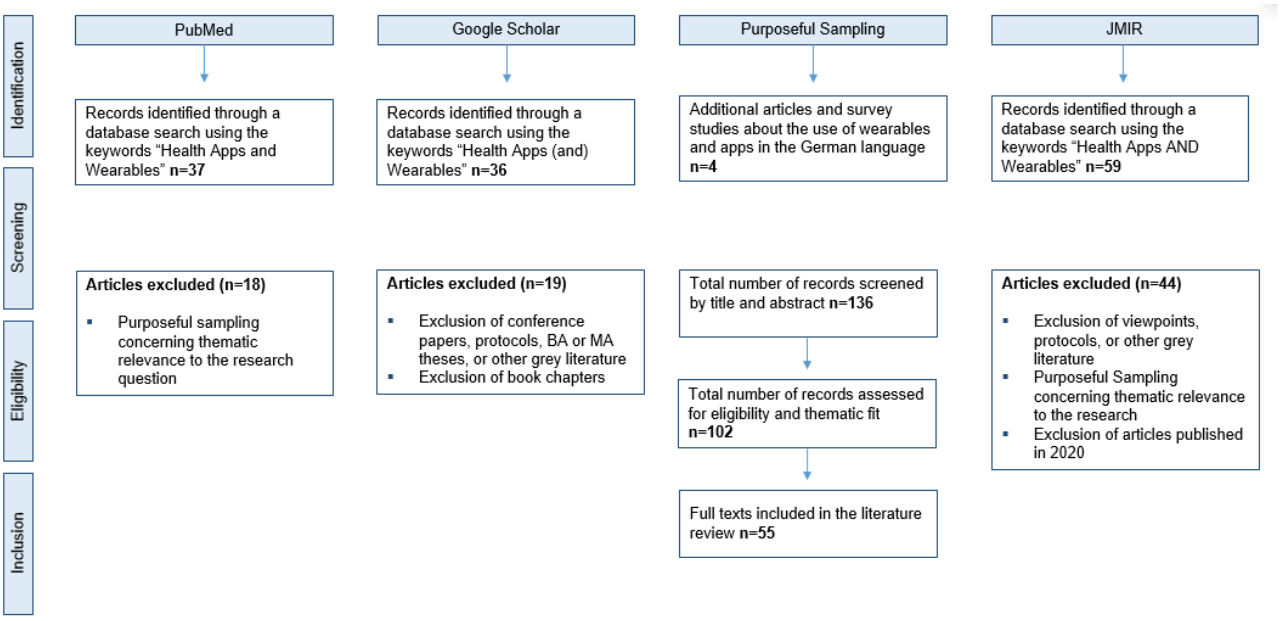
Methods

According to Munn et al [7], we conducted a scoping study to identify key concepts of the inclusion of health apps and wearables into the German statutory health care sector. The study aimed to draw a general picture about the risks and benefits of statutory financed mHealth solutions in Germany.

Scoping Method

We performed this study according to the guidelines of scoping studies by Colquhoun et al [8]. Colquhoun et al [8] advanced the 6 stages of scoping studies by Arksey and O'Melley [9]. They elaborated on different stages of research such as the identification of a research question and literature, study selection, charting data, summarizing, and consulting [9]. To ensure rigor and transparency, this literature review was guided by our research question [9]. We started the scoping study with a database search of PubMed using the keywords "health apps AND wearables" (Figure 1). The scoping of literature was limited to articles published in the English language from 2015 to March 2019 because literature on health apps and wearables, as well as the boom of using those technologies, experienced a steep increase in 2015 [10]. The search identified 37 potential items. A second search was conducted via Google Scholar by using the keywords "health apps (and) wearables," limiting the search again to literature published in the English language from 2015 to March 2019, and 36 items were identified. Then, another 2 articles in the German language and 2 survey studies in the German language were included in the study through purposeful sampling [11] because they were recommended. We conducted a third database search through JMIR using the search terms "health apps AND wearables" and identified 59 articles published from 2015 to March 2019 in English. We conducted other trial searches using other keywords such as "mHealth," "fitness apps," "health apps," and "health data sharing" but the sampled literature had little fit with the research question. Hence, when searching only for the search term "health apps," JMIR returned 698 search results. However, we chose the search term "health apps AND wearables" for our study because this is the closest that the published literature gets in terms of the German DIGA concept [12].

Figure 1. Scoping process for literature about health apps and wearables.



Identification of Relevant Articles

Conference papers, conference reports, protocols, viewpoints, letters, Bachelor and Master theses, or other grey literature were not included. First, we screened articles by title and abstract. Literature relating to the themes of patient treatment with health apps or wearables, preventive care with apps or wearables, market studies about health app and wearable use, data privacy concerns, and patient use of health apps and wearables were included in this study. Hence, duplicates and ineligibility were further reasons for exclusion. Regarding the inclusion themes selected via purposeful sampling [11], 55 of 136 articles were included in this scoping study, and we analyzed the articles using a qualitative thematic analysis approach (see Multimedia Appendix 1 and Multimedia Appendix 2).

Results

Of the 55 studies, 22 studies were literature, website, or app reviews; 16 studies were qualitative studies; and 17 studies were survey, interview, or quantitative studies. Most survey studies were not representative. Overall, we concluded that there is a growing amount of health app and wearable literature, but there is still room for additional research because not every aspect of the introduction of mHealth solutions into the health care system is known yet. There are few long-term studies on the effectivity of the use of health apps and wearables as a form of patient treatment. We have no insights about the effects of DIGA prescription and usage over 5, 10, or 20 years. Most articles we reviewed originated in Western Europe, the United States, and Canada.

After article scoping and conceptualization of the main findings, 4 main themes emerged: users of health apps and wearables; effectivity of health apps and wearables to improve health; the potential of bring-your-own, self-tracked data; and concerns and data privacy risks.

Users of Health Apps and Wearables

A study by GfK reported that 33% of survey participants from 16 different countries used wearables or health apps to track their fitness or health on average [6]. The main reasons for people to use these devices is to improve their personal level of fitness or for self-motivation. In Germany, about 28% of people currently track their health — more men than women and rather younger than older people [6]. Another survey conducted by Statista [13] showed similar results. Users mainly focus on self-optimization. The youngest user group (18-29 years) has the largest proportion of app users [13].

Wiesner et al [14] conducted a field study and surveyed participants from a regional road race event about their use of wearables. They decided to survey sport-enthusiast runners because they anticipated that mainly young and active people use health apps and wearables. The study showed that 73% of the runner community used one or more wearables to track their activity [14]. Just 1% of the respondents used wearables sponsored by their health insurer [14]. The authors further asked about data privacy concerns of nonvoluntary data sharing, and 42% of the respondents "stated that they would not be concerned if data were shared in such a manner" [14]. This result might be significantly different when surveying a group of chronically ill or nonactive people. The results further show that the willingness to share data with different agents decreases for respondents in older age groups [14]. Most respondents of a US market study used health apps and wearables to monitor personal activity, nutrition, weight loss, or learn a new exercise [15]. The majority of the surveyed users used their health or fitness apps at least once a day [15]. Just 20% of the respondents discovered an app through the recommendation of a physician [15]. Among the most frequent reasons for people to not use health apps and wearables were lack of interest, high prices, and lack of trust in data security [15].

Park et al [16] conducted a similar study in South Korea and achieved similar results. The main reasons to use health apps

and wearables were concerns about personal health status, self-optimization, innovative propensity, and trust in beneficial results. Surprisingly, the results indicated that the quality of the app has less influence on the decision whether to continue to use an app than social-cognitive factors [16]. Paré et al [17] also analyzed the motivation of people using health apps or wearables in Canada. They concluded that about 41% of the respondents used digital devices to self-track their health and physical activity (PA), which is significantly more than the German average. Furthermore, "a majority of digital self-trackers are young or mature adults (18-34 years), highly educated …, wealthy … and people who perceive themselves to be in good or very good health" [17]. Mosconi et al [18] and Ernsting et al [19] agreed with this statement and determined that young people in particular are interested in these technologies. Users feel generally more informed about their health when tracking different parameters, and 7 of 10 respondents improved or maintained their health condition by using an app or wearable [17]. Nevertheless, "one-third of consumer wearables end up in a drawer 6 months after purchase" [17]. This phenomenon occurs mainly with people with poor health or a chronic illness, indicating that this group loses interest in the technology when constantly reminded about a chronical condition or illness. Those people might feel pressured to be physically active [17].

Canhoto and Arp [20] agreed with Paré et al [17] by stating that many wearable and health app users stop using their devices after a while. Many insurance companies offer their members financial incentives and bonus programs to adopt a certain app or track specific health parameters [20]. The authors claimed that the inclusion of wearables and health apps in the health care system might have a significant positive influence on the treatment of chronic disease, like obesity or diabetes [20]. The widespread adoption and acceptance of these technologies are the key to their effectivity.

Christóvão [21] analyzed in his paper the influencing factors leading to app usage and the potential of health apps recommended and monitored by physicians. Perceived ease of use, perceived usefulness, peer influence, seniority, age, and gender were among the most important factors [21]. The author surveyed 199 fully qualified doctors and medical students to analyze the perceived usefulness of introducing health apps and wearables into patient care. Senior physicians and female physicians tended to use health apps less frequently if there was little peer influence, little perceived usefulness, and high complexity of usage [21]. A majority of the respondents could imagine recommending health apps and wearables to patients. Collado-Borrell et al [22], Davis et al [23], and Lipschitz et al [24] stated that many patients, nonetheless, already use health apps and wearables and are generally interested in the adaption of these technologies, independent of their age. However, Krebs and Duncan [15] rejected the view that all influencing factors are equally important. Wiesner et al [14] disagreed that gender significantly influences app usage, and Mackert et al [25] stated that health literacy plays an important role in the willingness to use these technologies.

Somers et al [26] conducted a contingent evaluation about the willingness-to-pay (WTP) for and willingness-to-accept (WTA) the use of health apps with different features. The results indicated that people value the promotion of wellbeing, social connectivity, and health care control [26]. Hence, Peng et al [27] identified the price of a wearable or health app as a significant influencing factor for the decision to adopt. The main reasons for people to abandon health apps or wearables after a certain period are, according to Peng et al [27], lack of time and effort and the lack of motivation and discipline. This means that apps or wearables alone cannot trigger a tremendous lifestyle change. The authors identified important reasons for people to use and continue to use health apps and wearables such as social competition, intangible rewards, tangible rewards, hedonic factors, and internal dedication [27]. To set incentives for nonactive or chronically ill patients to adopt health apps or wearables, tangible rewards like bonus programs might be the most promising tool in the future because "money is one of the biggest motivators" [27]. Petersen et al [28] concluded that tracking health parameters and communication through internet platforms triggers more self-consciousness and leads to patient empowerment.

Effectivity of Health Apps and Wearables to Improve Health

A study from the German Ministry of Health [29] assigned health apps and wearables a significant role in the future and singled out the importance of incorporating self-tracked data into the physician's daily routine and diagnostics. The stagnating telemedical development in Germany might be one of the major obstacles for the incorporation of DIGAs into the German health care system and needs further attention. However, Albrecht [29] argued that apps should be developed in cooperation with physicians, pharmaceutical companies, and health insurers to better meet the needs of the patients. The author claimed that the continuous use of health apps has a positive effect on personal health [29].

Mercer et al [30] conducted a participant's study and provided wearables to 32 chronically ill participants (aged >50 years), which they evaluated according to questions derived from the technology acceptance model. They found out that older and chronically ill people perceive wearables as "useful and acceptable." The use of wearables could enhance the level of PA because the devices create awareness of real motion [30]. Many older participants have not used a smartphone or tablet before and have strong concerns about their competencies. Yet, the technologies could remove barriers between physicians and patients [30]. Ehn et al [31] conducted a similar study. The authors concluded that the overall PA of the elderly increased during the study and that the wearables acted as a significant motivator [31]. However, they defined similar barriers for the use of wearables [31]. Schoeppe et al [32] reviewed 25 apps for children and adolescents and concluded that these apps have moderate quality overall. User engagement while using the app was not satisfactory, and the apps did not respond to individual needs. The authors ascribed to health apps for children and adolescents a high potential effectivity of sustainable behavioral change through gamification. They suggested, similar to Albrecht [29], cooperation of physicians, pharmaceutical companies, health insurers, and app developers [32]. Hartzler et al [33] and Hoffmann et al [34] also stressed the inclusion of

gamification and interactive features as main factors for the success of health apps and wearables.

Firth and Torous [35] concluded their literature search by stating that there is still little empirical research available on the effectivity of health apps, specifically as a complementary treatment for schizophrenia: "People with schizophrenia are willing and able to use smartphones to monitor their symptoms, engage in self-directed therapeutic interventions, and increase their physical exercise." Patients not officially diagnosed with schizophrenia or patients in acute stages report problems with app adherence [35]. Urrea et al [36] predicted that the use of health apps is an effective tool for the prevention of cardiovascular disease. Interventions via apps related to improvement and monitoring of smoking behavior, nutrition, and PA show positive results [36]. Hartmann et al [37], Christmann et al [38], and Ose et al [39] found significant potential of health apps and wearables for the treatment of depression. DIGAs might personalize care and reduce communication barriers with medical doctors. Gabriels and Moerenhout [40] and Martinez-Millana et al [41] concluded that the use of health apps and wearables help improve patients' awareness and health education.

The Potential of Bring-Your-Own, Self-Tracked Data

Haghi et al [42] ascribed to bring-your-own, self-tracked data an important role because of the predictions and simulations that could be achieved using big data: "The Internet of Things is a new concept, providing the possibility of health care monitoring using wearable devices." Health monitoring could be done to a large extent autonomously, using sensors like motion trackers, vital signs, and gas detectors [42]. Dimitrov [43] identified 4 main strategies: descriptive analysis, prescriptive analysis, predictive analysis, and simulations. The author predicted potential future savings in the health care sector because most patients could monitor their health by themselves and upload their data to a medical Internet of Things. Data analysis could be achieved using big data and digital health advisors, which could decrease the number of necessary annual visits to physicians [43]. Turankhia and Kaiser [44] agreed with Dimitrov [43] and identified the monitoring of patients at risk of atrial fibrillation with health apps and wearables as tools to increase the rate of early detection and therefore decrease physician visits. Heintzman [45] also argued that the management and monitoring of diabetes through apps could decrease costs for the health care system because the technologies offer individualized guidance. Henriksen et al [10] criticized that self-tracked data is, in most cases, uploaded to brand-specific repositories, which makes it difficult to share data with medical staff or compare data between different applications.

Vahabzadeh et al [46] identified mHealth primarily as a game changer in the treatment of depression and even as a measure of suicide prevention. The author stated that there is great potential to detect the risk of suicide early and to help individuals with specific apps tailored to their needs. However, medical doctors should not solely rely on these technologies for detection and treatment, given the tremendous pitfalls of a potential error [46]. Lüttke et al [47] agreed with the points made by Vahabzadeh et al [46]. They see great potential in the use of DIGAs as complementary to therapy.

Genes et al [48] researched the effectivity of asthma monitoring through health apps and concluded that there was improvement in asthma control and a decrease in necessary physician contact. More importantly, the use of the app helped to reduce barriers within patient-physician communication [48]. Yet, another study showed that the incorporation of bring-your-own, self-tracked data in preventive care programs might be very promising. The reason for the positive outlook is the advancement of patient education through data visualization and a better self-monitoring strategy [1]. However, the widespread adoption of these technologies and integration of the data in routine physician care are challenging [1]. Lobelo et al [1] recommended that health app developers, researchers, regulators, and medical staff conjointly develop solutions to ensure compliance, compatibility, and health data security. Brandt et al [49] conducted a study by interviewing general practitioners in Denmark, and a majority of the general practitioners already used health apps and are generally convinced about the effectiveness but do not "translate that into lifestyle change guidance for their patients." The authors suggested that health apps and wearables have significant potential to improve diagnostics and are a complimentary treatment for patients.

Chung et al [50] found that patients get better insights about their specific condition and feel empowered and connected. Cresswell et al [51] ascribed the integration of bring-your-own, self-tracked data into the daily routine of physicians as an aspirational role in preventive care and diagnostics. Furthermore, Cresswell et al [51] agreed with Chung et al [50] that self-monitoring of vital parameters and data visualization empower and educate patients.

Knight and Bidargaddi [52] concluded that self-management of mental diseases through apps leads to patient empowerment and the improvement of clinical care through better understanding. Ramkumar et al [53] agreed with this argument.

Concerns and Data Privacy Risks

Wichmann et al [54] criticized, despite all the potential benefits, the general academic enthusiasm about introducing DIGAs into the health care system, even though there is little empirical evidence about their long-term effectivity, or the usage over several years. Urban [55] conducted qualitative interviews to research the user perception of elderly people. The author claimed that health apps and wearables motivate elderly people to increase their activity, but they also cause them "to develop negative emotions that stand in a charged relationship to aging stereotypes." Elderly, who suffer from severe chronic conditions, feel discomfort integrating these technologies into their daily routine because the apps constantly remind them of their illness [55].

McCallum et al [56] agreed with Urban [55] and argued that the use of DIGAs are currently limited to mainly young and sportive people. To achieve widespread use, the usability and acceptability, especially of people with chronic conditions, need to be improved [56]. Data security issues are one of the main concerns for chronically ill people because they fear

discrimination in different parts of their daily life [56]. Montgomery et al [57] supported this claim and demanded government regulation to enhance fairness and equity but also to protect personal data from the sale to third parties.

Groß and Schmidt [58] suggested that patients could be overstrained with the amount of data and sensors available. Hence, patients are not sufficiently trained to read and properly analyze health data and peak graphs. They are not able to assess the data and identify their relevance, which could lead to misinterpretation [58]. The authors also listed positive effects resulting from the use of health apps and wearables for patients like efficiency, control, goal orientation, and better organization [58]. Another major problem discussed in the paper is the concern about data security, the consequences of potential data theft, and data sales to third parties [58].

Hicks et al [59] and Huckvale et al [60] discussed in their studies privacy risks that could result from the use of fitness and health apps. Users of health and fitness apps rely on the ethical operation of app services and need to trust the apps they use [59,60]. However, app services, especially those offering free operation, mainly sell the collected data to third parties and hide these conditions in very long policy terms. The authors examined the privacy policies of 79 popular health apps and found that 89% of the apps communicate with online services and 90% also communicate with "one or more third-party services directly" [60]. The authors criticized that most health and fitness apps "rely mainly on self-declared compliance" [60]. Armstrong [61] came to the same conclusion with a similar study and suggested government regulation for health data processing. Tabi et al [62] and Jamaladin et al [63] also criticized the lack of clarity of conventional app stores and emphasized the need for professional health app stores and medical doctors' recommendation to their patients. Becker et al [64] agreed with Huckvale et al [60] and Armstrong [61] because most health apps are not certified as a medical device, which means that their data protection terms are, in most cases, not supervised by a government agency. However, certification processes take a long time and are expensive. Incentives for the certification of apps are currently missing. However, the German digital care act enables fast track certification for DIGAs, which allows for early market access and a 1-year test phase to prove a positive health care effect [3].

Discussion

During the analysis of 55 of the 136 papers, we found 4 main themes or concepts regarding the introduction of DIGAs in the health care system: users of health apps and wearables; effectivity of health apps and wearables; the potential of bring-your-own, self-tracked data; and concerns and data privacy risks. In terms of the introduction of the digital care act in Germany, health apps and wearables are supposed to have an overall positive effect for patients. The literature shows that patients with chronic conditions especially could benefit from the DVG through self-monitoring and health education but also through reduced communication barriers with their physicians [29-31,35,43].

However, there is still a lack of long-term empirical evidence about the effect of statutory financed DIGAs. Yet, it is not very clear how health app and wearable developers should prove a positive effect on medical care for patients after their 1-year test-phase. Long test phases and costly control group trials are not feasible for health apps and wearables [5]. Many authors criticize the pure amount of health apps and wearables available on the market and the difficulty for people to choose one specific to their needs. They argue that integrating health care staff into the process of app development and recommendation and supervision by physicians would increase the potential benefits of the technology [1,10].

There are not just potential benefits but also severe direct and indirect privacy concerns and the fear of discrimination, for example, through the employer or health insurance company [65]. Users, especially in Germany, lack trust in many app providers concerning their data because of missing transparency. This is the reason why data privacy and data security are a major part of the DIGA certification process resulting from the digital care act. Hence, this is also why patient-tracked data is not automatically forwarded to the statutory sickness funds or the physicians. The patient should remain the owner of his data [5].

Transparency about data processing might be one of the major solutions to data privacy concerns. Users are generally more willing to share their data if application services are transparent about data processing than if it remains unclear or the user feels betrayed [66]. In European countries, personal data is understood to be personal property, and regulations such as the European General Data Protection Regulation (DSGVO) are set to protect this property [66].

In a second digitization phase, Germany could introduce another regulation that enables health care providers to offer patients a digital dividend to use their self-tracked data for research purposes. However, to price self-tracked health data might be very difficult because the users generally overestimate the price of their personal data: "By its nature, personal data is non-rival, cheap to produce, cheap to copy, and cheap to transmit" [66].

A recent study showed that many people in Germany are willing to share personal data in exchange for benefits or rewards: 12% agreed, 40% disagreed, and 48% did not want to answer the question [67]. Yet, 30 million German consumers use the Payback program initiated by the American Express Group, which involves selling consumer data for bonus points in certain stores [68]. Many people are not directly aware of the fact that they sell their data to Payback GmbH and the company sells the data to third parties [69]. When directly asked, people are often very sensitive to the commercial exploitation of personal data [70]. In the experiment by Cvrcek et al [70], the median bid accepted for location data was €43 (US $51.06). An experiment by Grossklags and Acquisti [71] showed that most participants are willing to sell their data but are not willing to pay for the protection. The average WTA for their data about individual quiz performance was US $7.06 and for their personal personal information was US $31.80. The WTP to protect both types of data was US $0.80 [71]. The authors discovered that the type of personal data is individual and emotionally charged, influencing the WTP and WTA decision. When participants

were asked about the number of their previous sexual partners, average WTA was US $2291.30, and WTP was US $12.10 [71]. Going a step further, when asked to auction their weight, age, and height, probands with a BMI below average demanded lower compensation to make their weight publicly available than probands with a BMI above average [72].

Hence, Von Wedel et al [73] showed that there is general interest in the inclusion of digital and mobile services in the German health care system. Yet, this gives a positive outlook for the success of the digital care act in Germany. According to the studies reviewed, we predict a high demand for DIGAs from young and healthy adults in the beginning. Yet, we believe that chronically ill and elderly patients benefit to a large extent from the regulation, which is why physicians and doctors should act as mediators and recommend, supervise, and accompany app use.

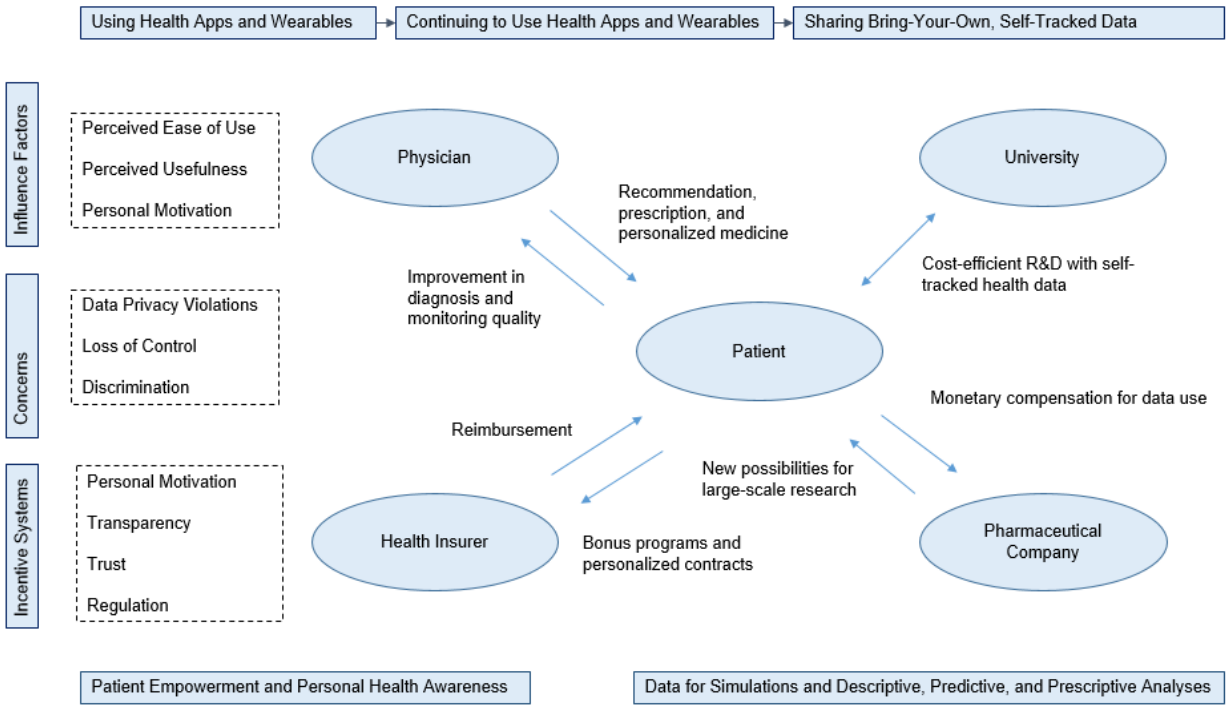
Three main stages of potential benefits for the German health care system were identified in the literature: usage of health apps and wearables [14,17,46], continuing usage of health apps and wearables [36,55], and sharing self-tracked data with agents in the health care sector [42,48]. Figure 2 shows the different stages mapped against the identified influence factors, concerns, and potential incentive systems.

The literature assigns each of the stages potential benefits when integrated into the health care system. The decision if individuals use health apps depends to a large extent on the perceived ease of use, perceived usefulness, trust, peer influence, personal health status, and technology literacy. Main concerns about the use of health apps and wearables discussed are data privacy violations or physical discomfort [15,16]. Whether an individual decides to continue to use a health app or wearable depends on the usefulness of the app to achieve certain goals, personal discipline, motivation, and trust. The concerns about continuing to use an app or wearable seem to be almost identical to the ones about starting to use an app, but even more sensitive to personal discomfort and the individual distortions of chronic diseases [21,27]. Presuming that the use of health apps and wearables has positive effects on the prevention of certain disease or aids treatments, the reasons why people stop using apps should be further studied, as well as potential incentive systems to assist people to continue to use these apps.

Some incentives named within the literature are bonus programs or physicians' recommendations. The last stage is the potential and willingness to bring along or share self-tracked data with different agents in the health care system. People seem to have very strong concerns about voluntarily sharing their self-tracked health data, which range from price discriminations to a lack of transparency and social embarrassment [26].

Referring to the research question of this paper, the digital care act and the introduction of statutory financed DIGAs could be considered societally beneficial. The widespread use of DIGAs allows patient empowerment, better monitoring of chronic diseases, and individualized advice. These benefits could not only reduce the number of mandatory visits to physicians and therefore the evergrowing expenses for the health care system but also lead to better resource allocation and improved treatment quality. Yet, Germany is the first country worldwide to introduce prescribed DIGAs. This is a significant chance to enhance digitization in the German health care sector and to build a foundation for a digital dividend to buy self-tracked patient data for research purposes. Yet, this experiment also bears risks when considering the volatile patient trust in data security.

Figure 2. Stages from use to continuous use to the sharing of self-tracked data.



Limitations

This study might be affected by the limited amount of available research resulting from the search terms. This might give a unilateral perspective on the effectivity of health apps and wearables. Hence, we are always concerned about the selection bias of articles. However, the multidisciplinary perspective on the field of study, enhanced through articles from different schools of thought and different research disciplines, as well as the applied rigor of scoping studies, have contributed to eliminate the selection bias to a large extent. Further research should be conducted after the first DIGAs are certified and have entered the German health care market.

Conclusions

To conclude, 55 of the 136 articles were analyzed within this scoping study. First, 4 key themes were identified: users of health apps and wearables; effectivity of health apps and wearables to improve health; potential of bring-your-own, self-tracked data; and concerns and data privacy risks.

In December 2019, Germany passed the digital care act, which enables the statutory financed prescription of digital health devices by medical doctors. Based on this scoping study, we predict an overall beneficial effect for German patients, treatment quality, and general health literacy of the population. The main benefits are going to be visible in the fields of preventive care and patient monitoring and disease management. Three main stages of potential benefits for the health care system were identified: using health apps and wearables, continuing to use health apps and wearables, and sharing bring-your-own, self-tracked data with different agents in the health care sector.

Acknowledgments

This research did not receive any grant or external funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study characteristics.
[DOCX File , 30 KB - mhealth_v8i9e16444_app1.docx]

Multimedia Appendix 2
Study aims and findings.
[DOCX File , 44 KB - mhealth_v8i9e16444_app2.docx]

References

1. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The Wild Wild West: A Framework to Integrate mHealth Software Applications and Wearables to Support Physical Activity Assessment, Counseling and Interventions for Cardiovascular Disease Risk Reduction. Prog Cardiovasc Dis 2016 May;58(6):584-594 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
2. Kotlikoff L, Hagist C. Who's Going Broke? Comparing Growth in Healthcare Costs in Ten OECD Countries. NBER Working Paper No. w11833 2005 Dec;188(1) [FREE Full text] [doi: 10.3386/w11833]
3. Der Bundestag. Bundesgesetzblatt Teil I Nr. 49. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale Versorgung Gesetz). 2019 Jul 22. URL: www.bgbl.de [accessed 2020-09-12]
4. Zeit Online. Jens Spahn: Krankenkassen Sollen Für Gesundheitsapps Zahlen. Die Zeit. 2019 May 15. URL: www.zeit.de/digital [accessed 2020-09-12]
5. Bundesinstitut für Arzneimittel und Medizinprodukte. Das Fast-Track-Verfahren Für Digitale Gesundheitsanwendungen (DiGA) Nach § 139e SGB V. DIGA Leitfaden. 2020. URL: www.bfarm.de/SharedDocs [accessed 2020-09-12]
6. GfK. Health and Fitness Tracking: Global GfK Survey. 2016 Sep. URL: https://cdn2.hubspot.net/hubfs/2405078/cms-pdfs/fileadmin/user_upload/country_one_pager/nl/documents/global-gfk-survey_health-fitness-monitoring_2016.pdf [accessed 2020-09-12]
7. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
8. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
9. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

10. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
11. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
12. Bundesministerium für Gesundheit. Verordnung über das Verfahren und die Anforderungen der Prüfung der Erstattungsfähigkeit digitaler Gesundheitsanwendungen in der gesetzlichen Krankenversicherung, 2020. Bundesgesetzblatt Teil I Nr. 18. 2020. URL: www.bgbl.de [accessed 2020-09-12]
13. Radtke R. Umfrage Zur Teilungsbereitschaft Von Gesundheitsdaten in Deutschland, 2016. Statista. 2018. URL: https://de.statista.com/statistik/daten/studie/712642/umfrage/umfrage-zur-teilungsbereitschaft-von-gesundheitsdaten-in-deutschland/ [accessed 2020-09-12]
14. Wiesner M, Zowalla R, Suleder J, Westers M, Pobiruchin M. Technology Adoption, Motivational Aspects, and Privacy Concerns of Wearables in the German Running Community: Field Study. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e201 [FREE Full text] [doi: 10.2196/mhealth.9623] [Medline: 30552085]
15. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
16. Park M, Yoo H, Kim J, Lee J. Why do young people use fitness apps? Cognitive characteristics and app quality. Electron Commer Res 2018 Jan 29;18(4):755-761. [doi: 10.1007/s10660-017-9282-7]
17. Paré G, Leaver C, Bourget C. Diffusion of the Digital Health Self-Tracking Movement in Canada: Results of a National Survey. J Med Internet Res 2018 May 02;20(5):e177 [FREE Full text] [doi: 10.2196/jmir.9388] [Medline: 29720359]
18. Mosconi P, Radrezza S, Lettieri E, Santoro E. Use of Health Apps and Wearable Devices: Survey Among Italian Associations for Patient Advocacy. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e10242 [FREE Full text] [doi: 10.2196/10242] [Medline: 30664455]
19. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons J, Kuhlmey A, Gellert P. Associations of Health App Use and Perceived Effectiveness in People With Cardiovascular Diseases and Diabetes: Population-Based Survey. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12179 [FREE Full text] [doi: 10.2196/12179] [Medline: 30920383]
20. Canhoto AI, Arp S. Exploring the factors that support adoption and sustained use of health and fitness wearables. Journal of Marketing Management 2016 Oct 26;33(1-2):32-60. [doi: 10.1080/0267257X.2016.1234505]
21. Veríssimo JMC. Usage intensity of mobile medical apps: A tale of two methods. Journal of Business Research 2018 Aug;89:442-447. [doi: 10.1016/j.jbusres.2017.12.026]
22. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology Patient Interest in the Use of New Technologies to Manage Their Disease: Cross-Sectional Survey. J Med Internet Res 2018 Oct 23;20(10):e11006 [FREE Full text] [doi: 10.2196/11006] [Medline: 30355554]
23. Davis TL, DiClemente R, Prietula M. Taking mHealth Forward: Examining the Core Characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]
24. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of Mobile Apps for Depression and Anxiety: Cross-Sectional Survey Study on Patient Interest and Barriers to Engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
25. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
26. Somers C, Grieve E, Lennon M, Bouamrane M, Mair FS, McIntosh E. Valuing Mobile Health: An Open-Ended Contingent Valuation Survey of a National Digital Health Program. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e3 [FREE Full text] [doi: 10.2196/mhealth.9990] [Medline: 30664488]
27. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
28. Petersen A, Schermuly AC, Anderson A. The shifting politics of patient activism: From bio-sociality to bio-digital citizenship. Health (London) 2019 Jul;23(4):478-494. [doi: 10.1177/1363459318815944] [Medline: 30526091]
29. Albrecht UV, editor. Opportunities and risks of health apps (CHARISMHA). Hanover, Germany: Peter L. Reichertz Institute for Medical Informatics at the TU Braunschweig and the Hannover Medical School; 2016:370.
30. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
31. Ehn M, Eriksson LC, Åkerberg N, Johansson A. Activity Monitors as Support for Older Persons' Physical Activity in Daily Life: Qualitative Study of the Users' Experiences. JMIR Mhealth Uhealth 2018 Feb 01;6(2):e34 [FREE Full text] [doi: 10.2196/mhealth.8345] [Medline: 29391342]

32. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
33. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth Design Space: A Mixed-Methods Analysis of Smokers' Perspectives. JMIR Mhealth Uhealth 2016 Aug 05;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
34. Hoffmann A, Christmann CA, Bleser G. Gamification in Stress Management Apps: A Critical App Review. JMIR Serious Games 2017 Jun 07;5(2):e13 [FREE Full text] [doi: 10.2196/games.7216] [Medline: 28592397]
35. 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]
36. Urrea B, Misra S, Plante TB, Kelli HM, Misra S, Blaha MJ, et al. Mobile Health Initiatives to Improve Outcomes in Primary Prevention of Cardiovascular Disease. Curr Treat Options Cardiovasc Med 2015 Dec;17(12):59. [doi: 10.1007/s11936-015-0417-7] [Medline: 26474892]
37. Christmann CA, Hoffmann A, Bleser G. Stress Management Apps With Regard to Emotion-Focused Coping and Behavior Change Techniques: A Content Analysis. JMIR Mhealth Uhealth 2017 Feb 23;5(2):e22 [FREE Full text] [doi: 10.2196/mhealth.6471] [Medline: 28232299]
38. Hartmann R, Sander C, Lorenz N, Böttger D, Hegerl U. Utilization of Patient-Generated Data Collected Through Mobile Devices: Insights From a Survey on Attitudes Toward Mobile Self-Monitoring and Self-Management Apps for Depression. JMIR Ment Health 2019 Apr 03;6(4):e11671 [FREE Full text] [doi: 10.2196/11671] [Medline: 30942693]
39. Ose SO, Færevik H, Kaasbøll J, Lindgren M, Thaulow K, Antonsen S, et al. Exploring the Potential for Use of Virtual Reality Technology in the Treatment of Severe Mental Illness Among Adults in Mid-Norway: Collaborative Research Between Clinicians and Researchers. JMIR Form Res 2019 Jun 10;3(2):e13633 [FREE Full text] [doi: 10.2196/13633] [Medline: 31199315]
40. Gabriels K, Moerenhout T. Exploring Entertainment Medicine and Professionalization of Self-Care: Interview Study Among Doctors on the Potential Effects of Digital Self-Tracking. J Med Internet Res 2018 Jan 12;20(1):e10 [FREE Full text] [doi: 10.2196/jmir.8040] [Medline: 29330140]
41. Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V. App Features for Type 1 Diabetes Support and Patient Empowerment: Systematic Literature Review and Benchmark Comparison. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e12237 [FREE Full text] [doi: 10.2196/12237] [Medline: 30463839]
42. Haghi M, Thurow K, Stoll R. Wearable Devices in Medical Internet of Things: Scientific Research and Commercially Available Devices. Healthc Inform Res 2017 Jan 21;23(1):4-15 [FREE Full text] [doi: 10.4258/hir.2017.23.1.4] [Medline: 28261526]
43. Dimitrov DV. Medical Internet of Things and Big Data in Healthcare. Healthc Inform Res 2016 Jul;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
44. Turakhia MP, Kaiser DW. Transforming the care of atrial fibrillation with mobile health. J Interv Card Electrophysiol 2016 Oct 15;47(1):45-50. [doi: 10.1007/s10840-016-0136-3] [Medline: 27306552]
45. Heintzman ND. A Digital Ecosystem of Diabetes Data and Technology: Services, Systems, and Tools Enabled by Wearables, Sensors, and Apps. J Diabetes Sci Technol 2015 Dec 20;10(1):35-41 [FREE Full text] [doi: 10.1177/1932296815622453] [Medline: 26685994]
46. Vahabzadeh A, Sahin N, Kalali A. Digital Suicide Prevention: Can Technology Become a Game-changer? Innov Clin Neurosci 2016;13(5-6):16-20 [FREE Full text] [Medline: 27800282]
47. Lüttke S, Hautzinger M, Fuhr K. [E-Health in diagnosis and therapy of mental disorders : Will therapists soon become superfluous?]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Mar 9;61(3):263-270. [doi: 10.1007/s00103-017-2684-9] [Medline: 29318339]
48. Genes N, Violante S, Cetrangol C, Rogers L, Schadt EE, Chan YY. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med 2018 Jun 20;1(1):23 [FREE Full text] [doi: 10.1038/s41746-018-0030-8] [Medline: 31304305]
49. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General Practitioners' Perspective on eHealth and Lifestyle Change: Qualitative Interview Study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88 [FREE Full text] [doi: 10.2196/mhealth.8988] [Medline: 29666045]
50. Chung AE, Sandler RS, Long MD, Ahrens S, Burris JL, Martin CF, et al. Harnessing person-generated health data to accelerate patient-centered outcomes research: the Crohn's and Colitis Foundation of America PCORnet Patient Powered Research Network (CCFA Partners). J Am Med Inform Assoc 2016 May 28;23(3):485-490 [FREE Full text] [doi: 10.1093/jamia/ocv191] [Medline: 26911821]
51. Cresswell KM, McKinstry B, Wolters M, Shah A, Sheikh A. Five key strategic priorities of integrating patient generated health data into United Kingdom electronic health records. J Innov Health Inform 2019 Jan 04;25(4):254-259 [FREE Full text] [doi: 10.14236/jhi.v25i4.1068] [Medline: 30672407]

52. Knight A, Bidargaddi N. Commonly available activity tracker apps and wearables as a mental health outcome indicator: A prospective observational cohort study among young adults with psychological distress. J Affect Disord 2018 Aug 15;236:31-36. [doi: 10.1016/j.jad.2018.04.099] [Medline: 29709718]
53. Ramkumar PN, Muschler GF, Spindler KP, Harris JD, McCulloch PC, Mont MA. Open mHealth Architecture: A Primer for Tomorrow's Orthopedic Surgeon and Introduction to Its Use in Lower Extremity Arthroplasty. J Arthroplasty 2017 Apr;32(4):1058-1062. [doi: 10.1016/j.arth.2016.11.019] [Medline: 27956125]
54. Wichmann F, Sill J, Hassenstein MJ, Zeeb H, Pischke CR. Apps zur Förderung von körperlicher Aktivität. Präv Gesundheitsf 2018 Oct 31;14(2):93-101. [doi: 10.1007/s11553-018-0678-6]
55. Urban M. 'This really takes it out of you!' The senses and emotions in digital health practices of the elderly. Digit Health 2017 Apr 12;3:2055207617701778 [FREE Full text] [doi: 10.1177/2055207617701778] [Medline: 29942589]
56. McCallum C, Rooksby J, Gray CM. Evaluating the Impact of Physical Activity Apps and Wearables: Interdisciplinary Review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
57. Montgomery K, Chester J, Kopp K. Health Wearables: Ensuring Fairness, Preventing Discrimination, and Promoting Equity in an Emerging Internet-of-Things Environment. Journal of Information Policy 2018;8:34. [doi: 10.5325/jinfopoli.8.2018.0034]
58. Groß D, Schmidt M. [Ethical perspectives on E-health and health apps : Is all that is achievable desirable?]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Mar 30;61(3):349-357. [doi: 10.1007/s00103-018-2697-z] [Medline: 29383392]
59. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019 Jun 3;2(1):45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
60. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 25;13(1):214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
61. Armstrong S. What happens to data gathered by health and wellness apps? BMJ 2016 Jun 23;353:i3406. [doi: 10.1136/bmj.i3406] [Medline: 27338775]
62. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
63. 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]
64. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
65. Wathieu L, Friedman AA. An Empirical Approach to Understanding Privacy Valuation. SSRN Journal 2007. [doi: 10.2139/ssrn.982593]
66. Spiekermann S, Acquisti A, Böhme R, Hui K. The challenges of personal data markets and privacy. Electron Markets 2015 Apr 29;25(2):161-167. [doi: 10.1007/s12525-015-0191-0]
67. GfK. Global GfK Survey: Willingness to Share Personal Data in Exchange for Benefits or Rewards. 2017. URL: https://cdn2.hubspot.net/hubfs/2405078/cms-pdfs/fileadmin/user_upload/country_one_pager/nl/images/global-gfk_onderzoek_-_delen_van_persoonlijke_data.pdf [accessed 2020-09-12]
68. Payback G. Daten & Fakten. Payback GmbH. 2019. URL: https://www.payback.net/at/ueber-payback/daten-fakten/ [accessed 2020-09-12]
69. Biermann K. Payback zahlt nichts zurück. Die Zeit. 2019 Mar 10. URL: https://www.zeit.de/online/2008/30/payback [accessed 2020-09-12]
70. Cvrcek D, Kumpost M, Matyas V, Danezis G. A Study on the Value of Location Privacy. 2006 Oct Presented at: 5th ACM Workshop on Privacy in Electronic Society - WPES '06; 2006; Alexandria, VA. [doi: 10.1145/1179601.1179621]
71. Acquisti A, Grossklags J. Privacy Attitudes and Privacy Behavior: Losses, Gains, and Hyperbolic Discounting. Economics of Information Security 2007:165-178. [doi: 10.1007/1-4020-8090-5_13]
72. Huberman B, Adar E, Fine L. Valuating Privacy. IEEE Secur. Privacy Mag 2005 Sep;3(5):22-25. [doi: 10.1109/msp.2005.137]
73. von Wedel P, Hagist C, Saunders K. Erratum: Die Digitalisierung der Arzt-Patienten Beziehung in Deutschland: Ein Discrete Choice Experiment zur Analyse der Patientenpräferenzen bezüglich digitaler Gesundheitsleistungen. Gesundh ökon Qual manag 2018 Feb 27;23(03):E1-E1. [doi: 10.1055/a-0580-9556]

Abbreviations

DIGA: digitale Gesundheitsanwendungen
DVG: Digitale Versorgung Gesetz
mHealth: mobile health

WTA: willingness-to-accept
WTP: willingness-to-pay

Edited by G Eysenbach; submitted 30.09.19; peer-reviewed by S Fetzer, L König; comments to author 24.04.20; revised version received 26.05.20; accepted 18.08.20; published 23.09.20.

Please cite as:
Heidel A, Hagist C
Potential Benefits and Risks Resulting From the Introduction of Health Apps and Wearables Into the German Statutory Health Care System: Scoping Review
JMIR Mhealth Uhealth 2020;8(9):e16444
URL: http://mhealth.jmir.org/2020/9/e16444/
doi:10.2196/16444
PMID:32965231

Original Paper

Mobile Breast Cancer e-Support Program for Chinese Women With Breast Cancer Undergoing Chemotherapy (Part 3): Secondary Data Analysis

Haihua Zhu[1*], BSc; Xiuwan Chen[1*], MSc; Jinqiu Yang[2*], PhD; Qiaoling Wu[3*], BSc; Jiemin Zhu[2], PhD; Sally Wai-Chi Chan[4], PhD

[1]Nursing Department, First Affiliated Hospital, Xiamen University, Xiamen, China

[2]Department of Nursing, School of Medicine, Xiamen University, Xiamen, China

[3]Hospital Infection Management Office, China-Japan Friendship Hospital, Beijing, China

[4]Global Engagement and Partnership Division, UON Singapore campus, University of Newcastle, Newcastle, Australia

[*]these authors contributed equally

Corresponding Author:
Jiemin Zhu, PhD
Department of Nursing, School of Medicine
Xiamen University
Room 222, Alice Lee building, School of Medicine, Xiamen University
Xiangan Nan Road, Xiangan District
Xiamen, 361102
China
Phone: 86 15960212649
Email: jieminzhu@xmu.edu.cn

Abstract

Background: Many app-based interventions targeting women with breast cancer have been developed and tested for effectiveness. However, information regarding the evaluation of the usage of these interventions is scarce. A better understanding of usage data is important to determine how women use apps and how these interventions affect health outcomes.

Objective: This study aimed to examine the usage duration and login frequency of an app-based intervention, the Breast Cancer e-Support (BCS) program, and to investigate the association between usage data and participants' demographic and medical characteristics.

Methods: This study is a secondary data analysis of a randomized controlled trial assessing the effectiveness of the BCS program. The BCS program contains four modules: Learning Forum, Discussion Forum, Ask-the-Expert Forum, and Your Story Forum. A total of 57 women in the intervention group accessed the BCS program during their 12-week chemotherapy. The app's background system tracked the usage duration and login frequency for each forum and the entire BCS program.

Results: The total usage duration per participant ranged from 0 to 9371 minutes, and the login frequency per participant ranged from 0 to 774 times. The Discussion Forum and the Learning Forum were the most frequently used modules. The general linear model showed that age, education, family monthly income, and employment were associated with BCS usage duration and/or login frequency. Age ($F_{1,45}$=10.09, P=.003, B=115.34, 95% CI 42.22-188.47) and education level ($F_{1,45}$=7.22, P=.01, B=1949.63, 95% CI 487.76-3411.50) were positively associated with the usage duration of the entire BCS program. Family monthly income was positively associated with the usage duration of the Learning Forum ($F_{1,45}$=11.85, P=.001, B=1488.55, 95% CI 617.58-2359.51) and the login frequency of the entire BCS program ($F_{1,45}$=4.47, P=.04, B=113.68, 95% CI 5.33-222.03). Employment was negatively associated with the usage duration of the Ask-the-expert Forum ($F_{1,45}$=4.50, P=.04, B=–971.87, 95% CI –1894.66 to –49.07) and the Your Story Forum ($F_{1,45}$=5.36, P=.03, B=–640.71, 95% CI –1198.30 to –83.11) and positively associated with the login frequency of the entire BCS program ($F_{1,45}$=10.86, P=.002, B=192.88, 95% CI 75.01-310.74). No statistical differences were found between BCS usage data and cancer stage, BMI, comorbidity, types of surgery, or cycles of chemotherapy.

Conclusions: Overall, this study found considerable variability in the usage of app-based interventions. When health care professionals incorporate app-based interventions into their routine care for women with breast cancer, the learning and discussion functions of apps should be strengthened to promote engagement. Additionally, characteristics of women with breast cancer,

such as age, level of education, income, and employment status, should be taken in consideration to develop tailored apps that address their particular needs and therefore improve their engagement with the app.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12616000639426; http://www.ANZCTR.org.au/ACTRN12616000639426.aspx

(JMIR Mhealth Uhealth 2020;8(9):e18896) doi:10.2196/18896

KEYWORDS

breast cancer; chemotherapy; mobile app; mHealth

Introduction

Chemotherapy often causes undesirable side effects in women with breast cancer, including fatigue, pain, sweating, swollen hands, and anxiety [1,2]. In China, access to adequate and continuous cancer care is challenging for women with breast cancer due to the increasing incidence of this type of cancer, shortage of specialized oncology health care professionals, and rising medical cost [3,4]. Innovative and easily accessible support is needed.

Connected Health, a health care delivery model to maximize health care resources, provides channels for participants to interact with clinicians and receive health-related services via new technology [5]. Mobile health (mHealth) is an example of this model of Connected Health [6]. Currently, thousands of health apps are available worldwide [7], with the most common type of cancer-related apps focussing on breast cancer [6].

We previously developed an app-based interactive program, called Breast Cancer e-Support (BCS) [8]. This program could improve women's self-efficacy and quality of life as well as reduce symptom interference during chemotherapy through 12-week access [9]. Women also perceived the BCS program to be beneficial in enriching knowledge, promoting confidence level, and improving emotional well-being [10]. Additionally, BCS usage duration was positively related to women's self-efficacy, social support, and quality of life during chemotherapy [9], which is consistent with a previous finding: "the more use, the better health outcome" [11].

A better understanding of app usage is critical to elucidate women's preferences for different forums and the ways the BCS program might affect health outcomes [12]. The usage statistics from an app are the real-world representations of individual usage processes [12], which are related to women's app perceptions and usage experience [13]. The evaluation of usage data might help to provide design recommendations to improve app engagement [13]. However, information on app usage, including duration and frequency, remains scarce [14-16]. Furthermore, it is unknown what modules are the most popular among participants.

Women with different characteristics, such as demographic and health related factors, are expected to have different usage patterns for app-based interventions [17]. Therefore, tailored interventions could be more effective and acceptable for women with breast cancer [18,19]. However, successful tailoring is challenging because of the heterogeneous characteristics of women with breast cancer [17,18,20,21]. Currently, findings are still inconclusive on the relationship between women's characteristics and their preference on app use [17,21-25]. For example, a recent systematic review revealed contrasting findings on whether age and baseline symptom status are associated with app usage [20]. A better understanding of the associations between women's characteristics and their usage patterns of the BCS program is critical to improve the design of app-based interventions.

The primary aim of this study was to examine the usage data of the BCS program by assessing the usage duration and login frequency in women randomized in an intervention group. The secondary aim was to investigate the associations between usage data and demographic and medical characteristics.

Methods

Participants

This study extracted data from a multicenter randomized control trial and focused on the analysis of 57 women randomly assigned to the BCS intervention group. All participants had to (1) be diagnosed with breast cancer, (2) be receiving chemotherapy, and (3) have access to the BCS program for 12 weeks. Women who had a concurrent serious physical illness or chronic mental illness were excluded from the study.

Intervention

The BCS app program, which could be used on both iPhone and Android mobile phones, aimed to provide information and social support to improve women's symptom management during chemotherapy. The intervention was conducted from May to November 2016. The full detailed description of the BCS program has been published [26].

The BCS program has four modules: Learning Forum, Discussion Forum, Ask-the-Expert Forum, and Your Story Forum. The Learning Forum offers evidence-based knowledge of breast cancer and management strategies for related symptoms. The Discussion Forum provides an anonymous interaction platform for women to communicate with peers and health care professionals. A seasoned health care professional moderates online discussion and provides expert advice when needed. The Ask-the-Expert Forum provides health consultation, where health care professionals respond to women's queries within 24 hours. The Your Story Forum presents videos of encouraging stories that include strategies for overcoming the challenges during chemotherapy [26]. A moderator facilitates the interaction between peers and health care professionals in the BCS program.

Demographic and Medical Characteristics

At baseline, women filled out paper questionnaires that collected demographic characteristics including age; marital status; education; employment; family income; BMI; and medical characteristics such as cancer stage, types of surgery, comorbidity, complications, and cycles of chemotherapy. Medical records were used to confirm the medical variables.

Usage Data: Usage Duration and Login Frequency

The usage data of the BCS program were measured as usage duration and login frequency; data from each forum and the entire BCS program were recorded for each individual for 12 weeks. Usage duration was defined as the time, in minutes, between login and logout for each forum and the entire BCS program. The operational definition of login frequency was the number of times that each woman logged into each forum and the entire BCS program during the 12-week intervention period. The app's background system tracked the usage data. If women were surfing on other functionalities of mobile phones or mobile phones were in standby mode, the app ran on background operational mode and the app background thread stopped calculating the usage data.

Statistical Analyses

All data analyses were performed using IBM SPSS 25.0. The mean (SD), medium (IQR), and maximum values were used to describe usage duration and login frequency. As the usage data were highly skewed, a general linear model was used to calculate the associations between usage data and demographic and medical characteristics. An α level of .05 of statistical significance was set for all analyses.

Results

Demographic and Medical Characteristics

A total of 57 women were included in this study. The mean age was 46.2 years (SD 8.5 years). All women were married, and 77% (44/57) were unemployed at the time of the study. Regarding educational level, 28% (16/57) had completed middle school, followed by elementary school (13/57, 23%), and high school (12/57, 21%). The monthly family income was in the range of US $149 to $738 for 60% (34/57) of the women and <US $148 for 25% (14/57) of them. Many of the women (28/57, 49%) were diagnosed with stage II breast cancer, followed by stage III breast cancer (19/57, 33%). A majority (45/57, 79%) had undergone mastectomy and only a few (3/57, 5%) had undergone breast conserving surgery.

BCS Usage Duration and Login Frequency

BCS usage data showed great variability. Only 4% (2/57) of the women did not log in the BCS program. For the whole BCS program, usage duration ranged from 0 to 9371 minutes, and frequency of logins varied from 0 to 774 times. The Discussion Forum and the Learning Forum were the most popular forums for women to log in and use. Table 1 shows the large difference between the mean and median of the BCS usage statistics.

Table 1. Usage duration and login frequency of Breast Cancer e-Support program for women with breast cancer receiving chemotherapy during the 12-week intervention (n=57).

Program and forums	Usage duration (minutes)			Login frequency (times)		
	Mean (SD)	Medium (IQR)	Maximum	Mean (SD)	Medium (IQR)	Maximum
Entire BCS[a] Program	1072.33 (2359.48)	100 (27-279)	9371	54.7 (131.4)	11 (5-27)	774
Learning Forum	399.51 (1139.95)	29 (12-162)	5926	13.75 (20.54)	6 (3-15)	95
Discussion Forum	412.09 (1364.85)	4 (1-43)	8539	53.89 (130.9)	7 (3-28)	715
Ask-the-Expert Forum	161.91 (1013.8)	2 (0-8)	7652	13.05 (25.26)	4 (2-12)	132
Your Story Forum	98.84 (597.73)	1 (0-13)	4480	5.18 (8.06)	3 (1-7)	52

[a] BCS: Breast Cancer Support.

Associations Between BCS Usage Data and Demographic and Medical Characteristics

Age, education, family monthly income, and employment were associated with BCS usage duration and/or login frequency. Age was positively associated with the usage duration of the entire BCS program ($F_{1,45}$=10.09, P=.003, B=115.34, 95% CI 42.22-188.47) and the Learning Forum ($F_{1,45}$=7.71, P=.008, B=49.93, 95% CI 13.70-86.13) (Table 2). Education level was found to be positively associated with the usage duration of the entire BCS program ($F_{1,45}$=7.22, P=.01, B=1949.63, 95% CI 487.76-3411.50) and the Discussion Forum ($F_{1,45}$=7.45, P=.01, B =1303.24, 95% CI 341.27-2265.21) (Table 2) as well as the login frequency of the Learning Forum ($F_{1,45}$=7.07, P=.01, B=17.82, 95% CI 4.32-31.32) and the Ask-the-expert Forum ($F_{1,45}$=6.17, P=.02, B=21.01, 95% CI 3.97-38.06) (Table 3). Family monthly income was positively associated with the usage duration of the Learning Forum ($F_{1,45}$=11.85, P=.001, B=1488.55, 95% CI 617.58-2359.51) (Table 2) and the login

frequency of the entire BCS program ($F_{1,45}$=4.47, P=.04, B=113.68, 95% CI 5.33-222.03), the Learning Forum ($F_{1,45}$=4.61, P=.04, B=17.31, 95% CI 1.07-33.56), and the Discussion Forum ($F_{1,45}$=6.68, P=.01, B=137.06, 95% CI 30.21-243.91) (Table 3). Employment status was negatively associated with the usage duration of the Ask-the-expert Forum ($F_{1,45}$=4.50, P=.04, B=–971.87, 95% CI –1894.66 to –49.07) and the Your Story Forum ($F_{1,45}$=5.36, P=.03, B=–640.71, 95% CI –1198.30 to –83.11) (Table 2) and positively associated with the login frequency of the entire BCS program ($F_{1,45}$=10.86, P=.002, B=192.88, 95% CI 75.01-310.74), the Learning Forum ($F_{1,45}$=11.76, P=.001, B=30.08, 95% CI 12.41-47.75), the Discussion Forum ($F_{1,45}$=10.31, P=.002, B=185.32, 95% CI 69.08-301.55), and the Your Story Forum ($F_{1,45}$=6.76, P=.01, B=9.30, 95% CI 2.09-16.51) (Table 3). No statistical differences were found between BCS usage data and cancer stage, BMI, comorbidity, types of surgery, or cycles of chemotherapy.

Table 2. Associations between usage duration in Breast Cancer e-Support program and demographic and clinical characteristics (n=57).

Variables	Usage duration									
	Entire BCS program		Learning Forum		Discussion Forum		Ask-the-Expert Forum		Your Story Forum	
	F test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
Age	10.09 (1,45)	.003[d]	7.71 (1,45)	.008[d]	2.46 (1,45)	.12	.70 (1,45)	.41	1.60 (1,45)	.21
Education[a]	7.22 (1,45)	.01[d]	2.17 (1,45)	.15	7.45 (1,45)	.01[d]	.14 (1,45)	.72	.003 (1,45)	.96
Family monthly income[b]	3.50 (1,45)	.07	11.85 (1,45)	.001[d]	.001 (1,45)	.98	.38 (1,45)	.54	2.68 (1,45)	.11
Cancer stage[c]	1.06 (1,45)	.31	.72 (1,45)	.40	3.30 (1,45)	.08	2.73 (1,45)	.11	.52 (1,45)	.47
BMI	.29 (1,45)	.60	.28 (1,45)	.60	.21 (1,45)	.65	.30 (1,45)	.59	1.08 (1,45)	.31
Employment	.04 (1,45)	.85	2.91 (1,45)	.10	.99 (1,45)	.32	4.50 (1,45)	.04[d]	5.36 (1,45)	.03[d]
Comorbidity	.44 (1,45)	.51	.001 (1,45)	.97	.12 (1,45)	.73	1.21 (1,45)	.28	1.72 (1,45)	.20
Types of surgery	.27 (2,45)	.76	.72 (2,45)	.49	.59 (2,45)	.56	.99 (2,45)	.38	.81 (2,45)	.45
Cycles of chemotherapy	.81 (2,45)	.45	2.76 (2,45)	.07	.56 (2,45)	.57	1.86 (2,45)	.17	.85 (2,45)	.44

[a]Regrouped into two categories: education “0=junior middle school and lower”, “1=higher school and higher.”

[b]Regrouped into two categories: family monthly income “0<USD 442”, “1≥USD 442.”

[c]Regrouped into two categories: “0=cancer stage I and II”, “1=cancer stage III or IV.”

[d]Significant at an α level of .05.

Table 3. Associations between login frequency in Breast Cancer eSupport program and demographic and clinical characteristics (n=57).

Variables	Login frequency									
	Login times: Entire BCS program		Login times: Learning Forum		Login times: Discussion Forum		Login times: Ask-the-Expert Forum		Login times: Your Story Forum	
	F test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
Age	.01 (1,45)	.91	1.39 (1,45)	.25	.51 (1,45)	.48	.36 (1,45)	.55	1.04 (1,45)	.31
Education[a]	3.61 (1,45)	.06	7.07 (1,45)	.01[d]	2.54 (1,45)	.12	6.17 (1,45)	.02[d]	2.19 (1,45)	.15
Family monthly income[b]	4.47 (1,45)	.04[d]	4.61 (1,45)	.04[d]	6.68 (1,45)	.01[d]	.04 (1,45)	.84	3.75 (1,45)	.06
Cancer stage[c]	.61 (1,45)	.44	1.71 (1,45)	.20	1.13 (1,45)	.29	1.69 (1,45)	.20	1.41 (1,45)	.24
Body mass index	.72 (1,45)	.40	.29 (1,45)	.59	.79 (1,45)	.38	2.11 (1,45)	.15	2.89 (1,45)	.10
Employment	10.86 (1,45)	.002[d]	11.76 (1,45)	.001[d]	10.31 (1,45)	.002[d]	3.66 (1,45)	.06	6.76 (1,45)	.01[d]
Comorbidity	1.48 (1,45)	.23	.95 (1,45)	.34	1.10 (1,45)	.30	.02 (1,45)	.90	.35 (1,45)	.57
Types of surgery	1.80 (2,45)	.18	1.48 (2,45)	.24	1.05 (2,45)	.36	.64 (2,45)	.53	1.58 (2,45)	.22
Cycles of chemotherapy	.50 (2,45)	.61	.43 (2,45)	.66	.35 (2,45)	.70	3.25 (2,45)	.05	.54 (2,45)	.59

[a]Regrouped into two categories: education “0=junior middle school and lower”, “1=higher school and higher.”

[b]Regrouped into two categories: family monthly income “0<USD 442”, “1≥USD 442.”

[c]Regrouped into two categories: “0=cancer stage I and II”, “1=cancer stage III or IV.”

[d]Significant at an α level of .05.

Discussion

Principal Findings

In this study, we found considerable variability in usage duration and frequency of the BCS program among women with breast cancer receiving chemotherapy. The large difference between median and mean usage data indicated usage polarization among the women. Additionally, we found that the Discussion Forum and the Learning Forum were the most popular forums. Age, education, family income, and employment were associated with women’s usage data.

Although BCS usage varied considerably, this study reported better usage duration and login frequency than a self-guided 4-month web-based intervention, called BREAst cancer ehealTH (BREATH), in which the total usage duration ranged from 0 to 2324 minutes (mean 337.2 minutes, SD 163.7 minutes) and the frequency of login ranged from 0 to 45 times (mean 11 times, SD 47 times) [27]. Moreover, our study had a lower proportion of non-users (2/57, 4%) than the BREATH intervention (7/70, 10%) [27]. BREATH, as a self-guided intervention, did not have professionals to moderate it. In contrast, this study involved a health care moderator who facilitated group interactions in the Discussion Forum and sent reminder messages to the corresponding doctors in the Ask-the-Expert Forum [9]. To explore whether a moderator is helpful in increasing participants’ engagement, further studies are warranted to compare usage statistics across different formats (self-guided versus moderated) within the same program.

We found usage polarization in the BCS program, which is consistent with the results of a study that explored the usage data of a web-based self-management intervention [27]. Usage polarization may reflect the particular preferences of each participant. In cases where reduced app engagement may undermine the potential effectiveness of the interventions [9], engaging strategies should be developed, such as self-tailored content, targeting participants who do not use the app as often as expected. However, tailoring the intervention to meet each individual’s needs presents multiple challenges [27]. Additionally, our process evaluation of the BCS program indicated that women’s poor physical and psychological conditions might hamper BCS usage [10]. Furthermore, in China, family members are often considered the primary caregivers of patients [28]. Therefore, to achieve better health outcomes for patients, family members should be involved in app-based interventions and information should be supplemented for them.

In this study, women used the Learning Forum and the Discussion Forum the most during their 12-week access to BCS. Research has found that patients with cancer use mHealth as an information resource and for emotional support [11,29,30]. The Learning Forum offered evidence-based information and strategies for managing symptoms, which helped women to better cope with breast cancer [26]. Thus, information support services may contribute to the usage of mHealth [30,31]. Meanwhile, the Discussion Forum provided a multi-interaction channel for in-depth discussion with peers and health professional moderators, which held particular appeal to these women [11,18,26,32,33]. Our findings concurred with the results of other studies that women tend not to use all the modules in apps, and instead prefer those that engage them [34]. Therefore, health care professionals should strengthen the key features of apps, such as learning and discussion functions, to enhance women's engagement.

Contrary to a previous study regarding eHealth use and age [35], we found that older women spent more time on the BCS program and the Learning Forum than younger women. In the abovementioned study, although older people were less likely to use eHealth intervention, the age disparity was found to be related to technological proficiency [35]. In this study, the user-centered design of the BCS program may help older women to alleviate their concern regarding technological ability [27]. Moreover, the interesting and relevant information gained from the Learning Forum may encourage older women to use eHealth more often [17]. Therefore, it would be beneficial to design user friendly apps and provide appropriate technical support and easy-to-understand information for older participants to meet their health information and care needs.

Consistent with some studies [20,21,23,34], our findings suggested that women with higher levels of education use the BCS program more frequently and spend more time on it. This may be interpreted as women with a higher education accepting new things more quickly and having better comprehension of the written information. These women also tend to spend more time learning health-related information [36]. Meanwhile, women with low education levels might experience difficulties in understanding the information, resulting in less interest in participating in app-based interventions [37]. However, owing to a lack of prior knowledge and health resources, women with low education levels may need to use the interventions to gain health-related information and consultation [17]. Health care professionals may improve the app design with lay language, more illustrations, and short videos to cater to the health information needs of women with different backgrounds to overcome the education barrier [10].

In this study, women with a higher family income used the Learning Forum and the Discussion Forum more than those with a lower family income, which is in agreement with other studies [17,23,37]. On the one hand, women with a low-income have less opportunities to access and use mobile devices at home and may experience stress when using the app-based program [23]. On the other hand, women with a higher family income might be able to afford spending more time to read and chat in the app-based program [17]. Besides, women with different family incomes may have different online surfing purposes. Women with a higher family income might go online to seek information, whereas women with a lower family income might seek entertainment [37]. However, low-income women with breast cancer have been reported to benefit more from web-based support because this support helps to fill the service gaps in the treatment of diseases and caters to their individual needs and preferences [38]. Thus, it is important for health care professionals to explore the needs of women with different income and improve the design and content of apps accordingly.

A previous study found that employed participants showed more interest in eHealth [35] and benefited more from eHealth [39]. Similarly, employed women used our BCS program more frequently. However, they spent less time in the Ask-the-Expert Forum and the Your-Story Forum, probably because participating in eHealth is time consuming and employed individuals have less leisure time [17]. In our process evaluation of the BCS program, women also suggested timely feedback from the experts as well as short and concise videos to convey information in these two forums [28]. These results indicate that modifying the design of eHealth to be timesaving may help facilitate usage engagement.

Strengths and Limitations

One strength of this study was the presentation of a method that directly and objectively captured the usage data of mobile apps. Thus, there was no reporting bias for mHealth intervention usage. Another strength was that this study analyzed the usage of each module in the program and its associations with women's demographic and medical characteristics. The findings could help researchers understand participants' preferences and explore strategies to improve app design.

As for the limitations, the study extracted usage data from a randomized control trial, and the sample size was small, which limited the generalizability of our results. Additionally, a previous study showed that the usage of apps changes over time [32]. Given the design of BCS, we could not track the dynamic usage data during the 12-week intervention. Longitudinal research with a larger sample size is warranted to gain deeper insight into the use of app-based interventions.

Conclusions

The insights gained from this study allow us to provide recommendations for further advances in the design and content of app-based health interventions. Overall, this study illustrated considerable variability in the usage of app-based interventions. When health care professionals incorporate app-based interventions into their routine care for women with breast cancer, the learning and discussion functions of apps should be strengthened to promote engagement. Additionally, characteristics of women with breast cancer, such as age, education level, income and employment status should be taken in consideration to develop tailored apps that address their particular needs and therefore improve their engagement with the app.

Acknowledgments

This study was supported by two grants (71503219 and 71974162) from the National Natural Science Foundation of China.

Authors' Contributions

HZ, XC, and JZ contributed to the conceptualization, data analysis, and methodology of the study. JZ contributed to funding acquisition. XC, JY, and QW contributed to original draft and writing. JZ and SC contributed to project administration, supervision, data validation, review, and editing. All authors have read and approved the manuscript.

Conflicts of Interest

None declared.

References

1. Uysal N, Toprak F, Kutlutsürkan S, Erenel A. Symptoms experienced and information needs of women receiving chemotherapy. Asia Pac J Oncol Nurs 2018;5(2):178-183 [FREE Full text] [doi: 10.4103/apjon.apjon_69_17] [Medline: 29607378]
2. Xu X, Cheng Q, Ou M, Li S, Xie C, Chen Y. Pain acceptance in cancer patients with chronic pain in Hunan, China: A qualitative study. Int J Nurs Sci 2019 Oct 10;6(4):385-391 [FREE Full text] [doi: 10.1016/j.ijnss.2019.09.011] [Medline: 31728390]
3. Goss PE, Strasser-Weippl K, Lee-Bychkovsky BL, Fan L, Li J, Chavarri-Guerra Y, et al. Challenges to effective cancer control in China, India, and Russia. Lancet Oncol 2014 Apr;15(5):489-538. [doi: 10.1016/S1470-2045(14)70029-4] [Medline: 24731404]
4. Bardhan IR, Thouin MF. Health information technology and its impact on the quality and cost of healthcare delivery. Decis Support Syst 2013 May;55(2):438-449. [doi: 10.1016/j.dss.2012.10.003]
5. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of breast cancer mobile applications: state of the practice review. Int J Med Inform 2018 Feb;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
6. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
7. Jahns RG. The market for mHealth app services will reach $26 billion by 2017. Research2guidance. 2013. URL: http://research2guidance.com/the-market-for-mhealth-app-services-will-reach-26-billion-by-2017/ [accessed 2020-08-30] [WebCite Cache ID http://research2guidance.com/the-market-for-mhealth-app-services-will-reach-26-billion-by-2017/]
8. Zhu J, Ebert L, Liu X, Chan SW. A mobile application of breast cancer e-support program versus routine care in the treatment of Chinese women with breast cancer undergoing chemotherapy: study protocol for a randomized controlled trial. BMC Cancer 2017 Apr 26;17(1):291 [FREE Full text] [doi: 10.1186/s12885-017-3276-7] [Medline: 28446141]
9. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
10. Zhu J, Ebert L, Guo D, Yang S, Han Q, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 1): qualitative study of women's perceptions. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e85 [FREE Full text] [doi: 10.2196/mhealth.9311] [Medline: 29643056]
11. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
12. Wang Z, Wang J, Maercker A. Program use and outcome change in a web-based trauma intervention: individual and social factors. J Med Internet Res 2016 Sep 09;18(9):e243 [FREE Full text] [doi: 10.2196/jmir.5839] [Medline: 27612932]
13. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related Internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
14. Quintiliani LM, Mann DM, Puputti M, Quinn E, Bowen DJ. Pilot and feasibility test of a mobile health-supported behavioral counseling intervention for weight management among breast cancer survivors. JMIR Cancer 2016 May 09;2(1):e4 [FREE Full text] [doi: 10.2196/cancer.5305] [Medline: 28410174]
15. Koneska E, Appelbe D, Williamson P, Dodd S. Usage metrics of web-based interventions evaluated in randomized controlled Trials: systematic review. J Med Internet Res 2020 Apr 16;22(4):e15474 [FREE Full text] [doi: 10.2196/15474] [Medline: 32297870]
16. Park YR, Lee Y, Kim JY, Kim J, Kim HR, Kim Y, et al. Managing patient-generated health data through mobile personal health records: analysis of usage data. JMIR Mhealth Uhealth 2018 May 09;6(4):e89 [FREE Full text] [doi: 10.2196/mhealth.9620] [Medline: 29631989]

17. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who follows eHealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. J Med Internet Res 2015 May 11;17(5):e115 [FREE Full text] [doi: 10.2196/jmir.3932] [Medline: 25963607]
18. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar 01;15(3):e34. [doi: 10.2196/jmir.2285] [Medline: 23454601]
19. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
20. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
21. Becker S, Brandl C, Meister S, Nagel E, Miron-Shatz T, Mitchell A, et al. Demographic and health related data of users of a mobile application to support drug adherence is associated with usage duration and intensity. PLoS One 2015 Jan;10(1):e0116980 [FREE Full text] [doi: 10.1371/journal.pone.0116980] [Medline: 25629939]
22. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013 Oct 16;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]
23. Arcury TA, Quandt SA, Sandberg JC, Miller DP, Latulipe C, Leng X, et al. Patient portal utilization among ethnically diverse low income older adults: observational study. JMIR Med Inform 2017 Dec 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
24. Mizuno M, Kataoka J, Oishi F. Relationship between the physical and psychosocial conditions of postoperative gastrointestinal cancer patients and their responses to an informational material. Asia Pac J Oncol Nurs 2017;4(1):53-60 [FREE Full text] [doi: 10.4103/2347-5625.199072] [Medline: 28217731]
25. Heiman H, Keinki C, Huebner J. EHealth literacy in patients with cancer and their usage of web-based information. J Cancer Res Clin Oncol 2018 Sep;144(9):1843-1850. [doi: 10.1007/s00432-018-2703-8] [Medline: 30006768]
26. Zhu J, Ebert L, Xue Z, Shen Q, Chan SW. Development of a mobile application of breast cancer e-support program for women with breast cancer undergoing chemotherapy. Technol Health Care 2017;25(2):377-382. [doi: 10.3233/THC-161292] [Medline: 28085020]
27. van den Berg SW, Peters EJ, Kraaijeveld JF, Gielissen MFM, Prins JB. Usage of a generic web-based self-management intervention for breast cancer survivors: substudy analysis of the BREATH trial. J Med Internet Res 2013 Aug 19;15(8):e170 [FREE Full text] [doi: 10.2196/jmir.2566] [Medline: 23958584]
28. Davidson PM, Daly J, Leung D, Ang E, Paull G, DiGiacomo M, et al. Health-seeking beliefs of cardiovascular patients: a qualitative study. Int J Nurs Stud 2011 Nov;48(11):1367-1375. [doi: 10.1016/j.ijnurstu.2011.02.021] [Medline: 21496819]
29. Yli-Uotila T, Rantanen A, Suominen T. Motives of cancer patients for using the Internet to seek social support. Eur J Cancer Care (Engl) 2013 Mar;22(2):261-271. [doi: 10.1111/ecc.12025] [Medline: 23320398]
30. Baker TB, Hawkins R, Pingree S, Roberts LJ, McDowell HE, Shaw BR, et al. Optimizing eHealth breast cancer interventions: which types of eHealth services are effective? Transl Behav Med 2011 Mar;1(1):134-145 [FREE Full text] [doi: 10.1007/s13142-010-0004-0] [Medline: 21709810]
31. Tarver WL, Haggstrom DA. The use of cancer-specific patient-centered technologies among underserved populations in the United States: systematic review. J Med Internet Res 2019 Apr 23;21(4):e10256 [FREE Full text] [doi: 10.2196/10256] [Medline: 31012855]
32. Han JY, Hawkins R, Baker T, Shah DV, Pingree S, Gustafson DH. How cancer patients use and benefit from an interactive cancer communication system. J Health Commun 2017 Oct;22(10):792-799 [FREE Full text] [doi: 10.1080/10810730.2017.1360413] [Medline: 28922091]
33. Barbabella F, Poli A, Hanson E, Andréasson F, Salzmann B, Döhner H, et al. Usage and usability of a web-based program for family caregivers of older people in three European countries: a mixed-methods evaluation. Comput Inform Nurs 2018 May;36(5):232-241. [doi: 10.1097/CIN.0000000000000422] [Medline: 29505433]
34. Couper MP, Alexander GL, Zhang N, Little RJA, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
35. Li P, Luo Y, Yu X, Wen J, Mason E, Li W, et al. Patients' perceptions of barriers and facilitators to the adoption of e-hospitals: cross-sectional study in Western China. J Med Internet Res 2020 Jun 11;22(6):e17221 [FREE Full text] [doi: 10.2196/17221] [Medline: 32525483]
36. Costa-Font J, Mossialos E, Rudisill C. When is the Internet a valued communication device for health information in Europe? Economics of Innovation and New Technology 2009 Jul;18(5):429-445. [doi: 10.1080/10438590802547159]
37. Močnik D, Širec K. The determinants of Internet use controlling for income level: cross-country empirical evidence. Information Economics and Policy 2010 Jul;22(3):243-256. [doi: 10.1016/j.infoecopol.2010.01.002]

38. Lu H, Shaw BR, Gustafson DH. Online health consultation: examining uses of an interactive cancer communication tool by low-income women with breast cancer. Int J Med Inform 2011 Jul;80(7):518-528 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.03.011] [Medline: 21530381]
39. Turnbull S, Cabral C, Hay A, Lucas PJ. Health equity in the effectiveness of web-based health interventions for the self-care of people with chronic health conditions: systematic review. J Med Internet Res 2020 Jun 05;22(6):e17849 [FREE Full text] [doi: 10.2196/17849] [Medline: 32459632]

Abbreviations

BCS: Breast Cancer e-Support
mHealth: mobile health

Edited by CL Parra-Calderón; submitted 26.03.20; peer-reviewed by D Attai, Y Pang, Q Sun; comments to author 14.06.20; revised version received 17.07.20; accepted 08.08.20; published 16.09.20.

Please cite as:
Zhu H, Chen X, Yang J, Wu Q, Zhu J, Chan SWC
Mobile Breast Cancer e-Support Program for Chinese Women With Breast Cancer Undergoing Chemotherapy (Part 3): Secondary Data Analysis
JMIR Mhealth Uhealth 2020;8(9):e18896
URL: http://mhealth.jmir.org/2020/9/e18896/
doi:10.2196/18896
PMID:32936087

Original Paper

One Drop App With an Activity Tracker for Adults With Type 1 Diabetes: Randomized Controlled Trial

Chandra Y Osborn[1,2], PhD, MPH; Ashley Hirsch[1], MA; Lindsay E Sears[1,3], PhD; Mark Heyman[1,4], PhD, CDE; Jennifer Raymond[5], MD, MCR; Brian Huddleston[1], JD; Jeff Dachis[1], MA

[1]Informed Data Systems Inc, New York, NY, United States

[2]Lirio, Nashville, TN, United States

[3]Sarah Cannon Research Institute, Nashville, TN, United States

[4]Department of Psychiatry, University of California San Diego, San Diego, CA, United States

[5]Division of Endocrinology, Department of Pediatrics, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Chandra Y Osborn, PhD, MPH
Lirio
901 Woodland Street
Nashville, TN, 37206
United States
Phone: 1 8604242858
Email: cosborn@lirio.co

Abstract

Background: In 2017, mobile app support for managing diabetes was available to 64% of the global population of adults with diabetes. One Drop's digital therapeutics solution includes an evidence-based mobile app with global reach, a Bluetooth-connected glucometer, and in-app coaching from Certified Diabetes Educators. Among people with type 1 diabetes and an estimated hemoglobin A_{1c} level≥7.5%, using One Drop for 3 months has been associated with an improved estimated hemoglobin A_{1c} level of 22.2 mg/dL (–0.80%). However, the added value of integrated activity trackers is unknown.

Objective: We conducted a pragmatic, remotely administered randomized controlled trial to evaluate One Drop with a new-to-market activity tracker against One Drop only on the 3-month hemoglobin A_{1c} level of adults with type 1 diabetes.

Methods: Social media advertisements and online newsletters were used to recruit adults (≥18 years old) diagnosed (≥1 year) with T1D, naïve to One Drop's full solution and the activity tracker, with a laboratory hemoglobin A_{1c} level≥7%. Participants (N=99) were randomized to receive One Drop and the activity tracker or One Drop only at the start of the study. The One Drop only group received the activity tracker at the end of the study. Multiple imputation, performed separately by group, was used to correct for missing data. Analysis of covariance models, controlling for baseline hemoglobin A_{1c}, were used to evaluate 3-month hemoglobin A_{1c} differences in intent-to-treat (ITT) and per protocol (PP) analyses.

Results: The enrolled sample (N=95) had a mean age of 41 (SD 11) years, was 73% female, 88% White, diagnosed for a mean of 20 (SD 11) years, and had a mean hemoglobin A_{1c} level of 8.4% (SD 1.2%); 11% of the participants did not complete follow up. Analysis of covariance assumptions were met for the ITT and PP models. In ITT analysis, participants in the One Drop and activity tracker condition had a significantly lower 3-month hemoglobin A_{1c} level (mean 7.9%, SD 0.60%, 95% CI 7.8-8.2) than that of the participants in the One Drop only condition (mean 8.4%, SD 0.62%, 95% CI 8.2-8.5). In PP analysis, participants in the One Drop and activity tracker condition also had a significantly lower 3-month hemoglobin A_{1c} level (mean 7.9%, SD 0.59%, 95% CI 7.7-8.1) than that of participants in the One Drop only condition (mean 8.2%, SD 0.58%, 95% CI 8.0-8.4).

Conclusions: Participants exposed to One Drop and the activity tracker for the 3-month study period had a significantly lower 3-month hemoglobin A_{1c} level compared to that of participants exposed to One Drop only during the same timeframe. One Drop and a tracker may work better together than alone in helping people with type 1 diabetes.

Trial Registration: ClinicalTrials.gov NCT03459573; https://clinicaltrials.gov/ct2/show/NCT03459573.

(JMIR Mhealth Uhealth 2020;8(9):e16745) doi:10.2196/16745

KEYWORDS

diabetes; type 1 diabetes; digital therapy; mobile app; coaching; glucometer; activity tracker

Introduction

Diabetes is common, costly, and can have serious consequences. An estimated 30 million people in the United States are living with diabetes, 1.3 million of whom have type 1 diabetes (T1D) [1] indicated by the pancreas producing little or no insulin. Approximately US $327 billion is spent annually to treat diabetes, complications from diabetes, and employees' losses in productivity [2]. Moreover, at least 11% of annual US deaths can be attributed to a diabetes complication [3]. A hemoglobin A_{1c} level of <7% reduces the risk of developing diabetes complications [4,5], but can be hard to achieve for a variety of reasons [6]. People with T1D achieve "at goal" blood glucose with a combination of insulin therapy, carbohydrate monitoring, blood glucose monitoring, and physical activity [7,8]. Frequent, painful insulin injections and finger pricks have historically made it difficult to administer insulin and check blood glucose as recommended [9]. However, digital advances in the last 20 years have made administering insulin (ie, via insulin pumps) and monitoring blood glucose (ie, via continuous glucose monitors) much easier [10,11]. Other consumer technologies such as health apps, digital therapies, and wearable activity trackers can aid in the management of T1D, making it easier and more convenient to perform and monitor self-care activities [12].

Consumer apps and activity trackers can help people with T1D meet their glycemic targets [13-16]. Among adolescents with T1D, using an activity tracker has been associated with being more physically active and having an improved average time in-range blood glucose level after a 3-month period [13]. Among children with T1D wearing activity trackers with health care providers remotely monitoring their tracker data (and other data) [14], quality of life and hemoglobin A_{1c} improved after 3 months [14]. However, 3-month hemoglobin A_{1c} benefits among adults with T1D using trackers independent of remote monitoring is unknown.

Adults with T1D can use the One Drop mobile smartphone app with or without activity trackers (eg, Apple Watch [17,18]) [16]. One Drop's app reads and displays activity data from trackers and other devices and is rated among the top three diabetes apps in the world [19]. The One Drop Chrome Bluetooth-connected meter syncs and displays blood glucose readings in the app. One Drop's Certified Diabetes Educators (CDE "coaches") remotely monitor user data and offer in-app education, strategies, and support. In observational studies, people with T1D or type 2 diabetes using One Drop's app on Apple Watch averaged a –1.2% to –1.3% absolute estimated hemoglobin A_{1c} improvement [17,18].

Studies consistently associate using One Drop's solution with improved estimated hemoglobin A_{1c}, but none of these studies used a randomized controlled trial design or included people with T1D using an activity tracker or smartwatch. Therefore, we conducted a prospective randomized controlled trial with adults with T1D to evaluate the 3-month effect of using One Drop and a new-to-market activity tracker on hemoglobin A_{1c}.

Methods

Study Design

Solutions IRB, a private Institutional Review Board (IRB) registered with HHS #IRB00008523 and accredited by the Association of Human Research Protection Programs, approved all study procedures prior to recruiting participants. The study design was a pragmatic, parallel group, randomized controlled trial. Study personnel used a block randomization scheme of 100 groups of two randomization blocks to randomize participants to one of two conditions: (1) One Drop's digital therapeutics solution (ie, the mobile app, in-app coaching, Bluetooth-connected meter with a 3-month supply of test strips) and an activity tracker at the start of the intervention period or (2) One Drop at the start of the intervention period and an activity tracker after completing 3-month follow-up measures. Participants and study personnel were unblinded to the condition assignment. Study personnel did not tell participants which condition was the intervention of interest and which one was the comparator, but participants may have inferred this on their own. Study instructions, consent, Health Insurance Portability and Accountability Act (HIPAA) authorization forms, and self-reported surveys were self-administered online using HIPAA-compliant surveys and forms. Participants used a mail-in hemoglobin A_{1c} test to self-collect and supply two blood specimens. Study personnel provided virtually disseminated instruction and support (via phone and email) to remotely eligible participants in their respective study conditions. Only participants accessing all intervention components were considered to be enrolled in the trial.

Recruitment

Facebook advertisements and One Drop's email list of noncustomers (ie, people never having used a One Drop meter, testing supplies, or coach) remotely recruited potential participants from March through May of 2018. Online advertisements and email messages briefly described study eligibility (eg, diagnosis of T1D), study scope (eg, 3-month duration), and asked people interested in the study to click a link to obtain in-depth information about the study and complete an online, HIPAA-compliant survey to self-screen for initial eligibility.

Eligibility

Initially eligible individuals met screening survey criteria. They self-reported an age of 18-75 years, had a valid US mailing address, a diagnosis of T1D for ≥1 year, were not currently participating in a diabetes education or coaching program, were not pregnant or planning to become pregnant, were using an Android or iOS smartphone, and had never used the activity tracker or One Drop (no app activity, 7-day trial, testing supply subscription, or coaching). An application programming interface (ie, a software intermediary for transferring data from

one application [One Drop's database] to another [an Excel spreadsheet]) automatically and objectively checked whether respondents had previously used any aspect of One Drop. Any participant not meeting initial eligibility criteria was notified of this on the screening survey's final landing page and thanked for their interest in the study.

People who self-screened as eligible followed a different path. They landed on an electronic IRB-approved consent form and HIPAA authorization form, requiring review and signature. All respondents were invited to contact study personnel to receive a verbal explanation of the forms or have any study-related questions answered. Once respondents electronically signed both forms, they landed on an online, HIPAA-compliant baseline survey. Upon completing the baseline survey, DTI Laboratories, Inc. shipped an AccuBase hemoglobin A_{1c} test kit to each respondent's mailing address. Study personnel provided written, illustrated, and video hemoglobin A_{1c} test kit instructions and offered over-the-phone help in collecting a blood sample. The participants returned blood samples to the lab in a preaddressed and prestamped box. The lab processed each sample and uploaded results into a HIPAA-compliant online portal. Study personal reviewed each result to determine hemoglobin A_{1c} eligibility.

Participants with a hemoglobin A_{1c} level≥7% were considered eligible for the study, randomized to one of the two conditions, and notified of their hemoglobin A_{1c} test result and condition assignment. People deemed ineligible (ie, hemoglobin A_{1c}<7%) were also notified of their hemoglobin A_{1c} test result, told they did not meet the hemoglobin A_{1c} criterion for participation, and were thanked for their interest in the study.

Data Collection and Procedures

Baseline Data Collection

The baseline survey collected demographic and diabetes information and responses to other self-report measures. The hemoglobin A_{1c} test determining study eligibility also served as the participants' measure of baseline hemoglobin A_{1c}.

Randomization

We used an online randomizer to block-randomize 100 groups of two randomization blocks to randomize participants to receive One Drop's digital therapeutics solution and an activity tracker at the start of the intervention period or One Drop at the start of the intervention period and an activity tracker after completing follow-up measures.

One Drop

The digital therapeutics solution includes the accurate Food and Drug Administration (FDA)-approved One Drop Chrome Bluetooth-connected glucometer [20] and testing supplies, One Drop mobile smartphone and smartwatch app, and One Drop coaching programs. One Drop coaches are real-life CDEs providing the first digitally delivered diabetes education program accredited by the American Diabetes Association. Users of One Drop's digital therapeutics solution have 24/7 in-app access to their personal CDE coach who answers questions, offers tips and advice, and provides practical and emotional support and accountability for daily self-care. One Drop's evidenced-based app [21] is available on iOS, Android, watchOS, and Amazon's Alexa and has been downloaded in every country in the world. Features include reminders to perform and track self-care, a "Community" section to bolster normative support, and education and skills training via the dynamic "Newsfeed" section along with the coaching chat section and programming content. Data reports can be viewed in the app, printed, and emailed.

As is the case with all apps, occasional minor bug fixes are typical; however, none of these resulted in major system failures or downtimes during the study period.

Activity Tracker

The wrist-worn device tracks activity and swimming, monitors heart rate, includes a built-in GPS, real-time statistics (eg, pace and distance), phone-free music to exercise with, and personalized workouts. With each workout, the software learns about a user's fitness level, makes personalized recommendations, and gives dynamic feedback. Third-party app developers such as One Drop can make device-compatible apps. One Drop's app on the device is an at-a-glance display of the last minutes of activity, grams of carbohydrates last consumed, last blood glucose reading, and last medications taken.

Once randomized to a condition, eligible participants received an email message containing their condition assignment, a series of instructions, and a unique verification code. The email message instructed participants to first download the One Drop mobile app on iOS or Android with embedded links to both formats for direct access. Next, participants were instructed to open the One Drop app, create an account, and enter their unique coaching verification code (from the email). Finally, participants were given a link to One Drop's online store and instructed to trigger a no-cost shipment of the activity tracker, One Drop meter, and testing supplies for the 3-month study period. When needed, study personnel assisted participants with completing these steps via a phone call or email exchange.

Attempts were made to keep study personnel blinded to condition. Randomization and condition assignment occurred separately with different researchers. A single researcher responded to participants' technical and research-related questions. All study procedures were digitally accessible. Most procedures were automated and self-administered, maintaining limited researcher touch and balance between groups. Furthermore, researchers and participants were separated the vast majority of the time.

One Drop and Activity Tracker at Study Start

Participants assigned to the One Drop and activity tracker condition were exposed to One Drop's comprehensive digital therapeutics solution and an integrated activity tracker at the start of the study. First, participants were mailed the activity tracker, One Drop meter, and testing supplies. Participants were also emailed "how-to" videos, written instructions, and offered study personnel support for setting up their devices and adjusting their smartphone settings to view activity data in the One Drop app. How-to videos included how to set up their activity tracker

on a smartphone or computer, automatically record exercises, sync data with other health apps, personalize notifications, and perform other customizations. Written instructions were supplied on how to directly integrate the tracker with One Drop to view tracker activity data (minutes and steps) in the One Drop app and view One Drop data on the tracker's clockface. Study personnel were also available 24/7 to answer participants' technical questions via email, text message, and phone calls.

Upon receipt of the activity tracker and One Drop's meter/strips, participants were instructed to connect with their coach (a real-life CDE) in the One Drop app, download the tracker-compatible One Drop app, and link their One Drop and activity tracker accounts. Finally, participants were instructed to use One Drop's app on their smartphone and activity tracker, One Drop's meter/strips, and in-app coaching "as needed" for the 3-month study period. Participants could initiate two-way communication with their coach about a wide range of diabetes self-care topics. For physical activity, topics may include, but are not limited to, those related to a participant's tracked activity (visible on the coach's dashboard), reasonable goal setting around minutes of activity or steps walked per day/week, or how to manage blood glucose levels before and after a bout of exercise.

One Drop and Activity Tracker at Study End

Participants assigned to the One Drop only condition were mailed the One Drop meter and testing supplies. The activity tracker was shipped after completing the follow-up survey and hemoglobin A_{1c} test. Participants connected with their One Drop coach via the One Drop app, and were instructed to use the app, meter, and in-app coach "as needed" for the 3-month study period.

Follow-Up Data Collection

After 3 months, participants in both conditions received an initial email and then a series of reminder emails instructing them to complete an online, HIPAA-compliant follow-up survey hyperlinked in the email. After participants completed this survey, DTI Laboratories, Inc. mailed a hemoglobin A_{1c} test kit to assess participants' 3-month hemoglobin A_{1c}. Again, study personnel sent instructions in various formats along with study contact information to aid with collecting a blood sample. Participants returned blood samples in a preaddressed and prestamped box. The lab processed each sample and uploaded results into the HIPAA-compliant online portal.

Study personal reviewed follow-up hemoglobin A_{1c} results, shared results with each participant, thanked them for their participation, and sent a discount code for a monthly and annual One Drop subscription. After participants in the One Drop only condition completed the follow-up survey and hemoglobin A_{1c} test, study personnel shipped their activity tracker to their mailing address. There were no methodological changes during the study period.

Compensation

Participant compensation included 3 free months of One Drop testing supplies and in-app coaching, and a free One Drop Chrome Bluetooth-connected meter and activity tracker to keep beyond the study period.

Measures

Demographic Characteristics

The baseline survey collected self-reported age, gender, race/ethnicity, education, annual income, and health insurance status.

Health Status

Health status information included self-reported number of years since a diabetes diagnosis and BMI.

Digital Health History

At baseline, participants self-reported whether or not they had ever used a blood glucose monitoring device (finger stick, continuous glucose monitor, flash monitor), a diabetes app, the new-to-market wearable tracker being used in the study, or any other wearable tracker to manage their health. Response options were "yes" or "no."

Digital Usability

At follow up, participants self-reported on a 7-point Likert scale ranging from "extremely hard" to "extremely easy" how hard to easy it was to use One Drop's meter, diabetes app, and in-app coaching. Participants in the One drop and activity tracker condition also self-reported how "extremely hard" to "extremely easy" it was to use the new-to-market activity tracker included in the study.

Digital Engagement and Attrition

At follow up, participants self-reported on a 7-point Likert scale ranging from "never" to "always" how often they used One Drop's meter, diabetes app, and in-app coaching when needing to check blood glucose, manage diabetes, or get help with diabetes, respectively. We then verified self-reported One Drop engagement with objectively collected data through the One Drop app. Participants in the One Drop and activity tracker condition also self-reported how often they used the new-to-market activity tracker during the 3-month study period.

Hypoglycemic Events

Hypoglycemia (ie, an extremely low blood glucose level requiring assistance) is a rate-limiting factor in the management of T1D and optimization of blood glucose. We accounted for the occurrence of hypoglycemia during the study period by asking all participants a single question in the 3-month follow-up survey: "In the past 3 months, how many times have you had a low blood sugar requiring help?" Responses were based on counts.

Glycated Hemoglobin A_{1c}

Hemoglobin A_{1c} was measured twice. Self-administered AccuBase A_{1c} Mail-In Test Kits (DTI Laboratories, Inc., Thomasville, GA, USA) were used to assess baseline and 3-month hemoglobin A_{1c} levels. This test is FDA-approved, certified by the National Glycohemoglobin Standardization Program, Clinical Laboratory Improvement Amendments-waived, and a highly accurate assessment of

hemoglobin A_{1c} used in randomized and nonrandomized trials [22,23]. It is a nonfasting, finger stick, whole blood mail-in test. Upon supplying a blood sample, specimens are processed at a central lab.

Statistical Analyses

Statistical analyses followed the Consolidated Standards of Reporting Trials (CONSORT) guidelines for randomized trials. Data were analyzed using SPSS version 24. Sample data at baseline overall and separately by group are described as means (SD) or counts (n, %) as appropriate. We followed CONSORT guidelines for testing and treatment of baseline group differences [24].

Counts characterized any engagement with One Drop during the study period, and for completing the study overall and by condition assignment. Two Chi-square tests were used to assess One Drop engagement and to assess study completion differences between groups.

Descriptive statistics were also used to characterize baseline digital health history, follow-up digital usability and engagement, study attrition, and the number of hypoglycemic events experienced during the study period overall and by study condition. Chi-square tests and Mann Whitney *U* tests were used to assess these particular baseline and follow-up group differences.

Multiple imputation was used to correct for missing data [25] on income (n=4) and follow-up hemoglobin A_{1c} (n=10). In both groups, variables used to impute included nonmissing age, gender race/ethnicity, education, insurance status, diabetes duration, baseline BMI and hemoglobin A_{1c}, and available data on income. Data were imputed separately by study condition. Imputed data were constrained by condition-specific minimum and maximum values. There were 20 imputations per condition. Data were merged prior to conducting intent-to-treat (ITT) and per protocol (PP) analyses.

Analysis of covariance (ANCOVA) models [26] were used to test the group effect on follow-up hemoglobin A_{1c} controlling for baseline hemoglobin A_{1c}. For noncrossover, parallel group randomized controlled trials, CONSORT guidelines recommend reporting results from ITT and PP analyses [27,28]. ITT analysis preserves baseline condition assignment and avoids overestimating group effects [29]. In contrast, PP analysis may exaggerate group effects by including only participants receiving the allocated intervention and completing the study as intended [30,31]. In pragmatic trials, the appropriate reporting of both results can aid with scientific and clinical interpretation [32]. We examined ANCOVA assumptions before conducting ANCOVA models testing 3-month hemoglobin A_{1c} group differences.

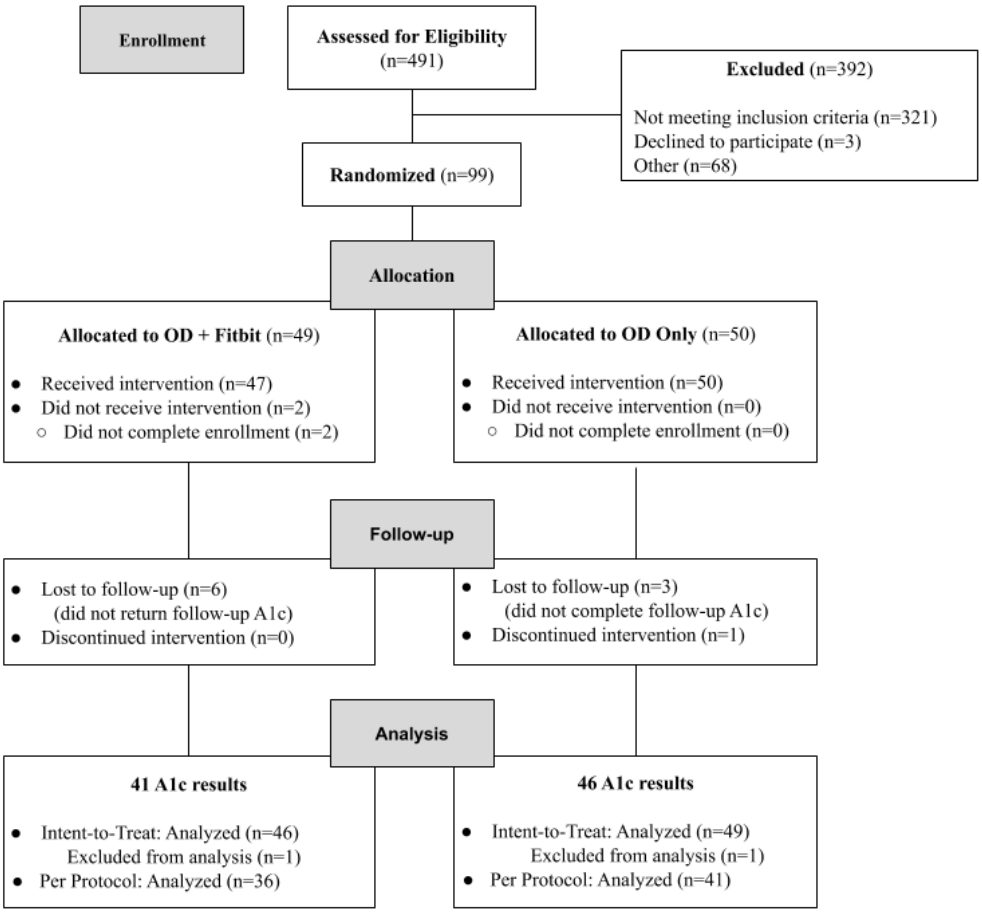
Results

Participants

Recruitment and enrollment occurred from March through May of 2018, and the last follow up was in August 2018. As shown in the CONSORT diagram (see Figure 1), 491 people self-screened for initial eligibility; 129 screened as initially eligible, completed the informed consent, HIPAA authorization, and the baseline survey, and were shipped a hemoglobin A_{1c} test kit. A total of 112 people returned the kit with a blood sample; 99 people satisfied the eligibility criterion of hemoglobin A_{1c}≥7% and were randomized to the One Drop and activity tracker condition or One Drop only condition, 97 of whom received the intervention, defined as creating a One Drop account and enrolling in One Drop coaching during the study period. Study participation resulted in no reported harm, unintended effects, or adverse events. There were also no privacy breaches, severe technical problems, or unexpected/unintended incidents during the study.

Quality assurance efforts identified two participants ineligible for the study (ie, one participant in each condition had used One Drop before) who were excluded from all analyses, resulting in 95 participants for ITT analysis and 77 participants for PP analysis. Participants in the PP analysis used One Drop (app, meter, and/or coaching) at least once during the study period and provided a follow-up hemoglobin A_{1c} blood specimen.

Figure 1. Study flow diagram. OD: One Drop; A_{1c}: hemoglobin A_{1c}.



Digital Health History

All eligible participants self-reported current smartphone utilization. At baseline, all enrolled participants self-reported having previously used a blood glucose monitoring device (eg, finger stick, continuous glucose monitor, flash device), but also reported that they had never used a diabetes app to manage their health or the new-to-market tracker used in the study. However, 61% (58/95) had used some type of wearable device to manage their health prior to the study. Prior experience with a wearable health device did not differ between groups (χ^2_1=0.77, *P*=.38).

Among participants that reported on the ease of use or usability of One Drop at follow up, 98% (88/90) reported One Drop's meter was easy to use, 83% (75/90) reported the app was easy to use, and 75% (67/89) reported in-app coaching was easy to use. The usability of One Drop was comparable between groups (*P*=.49 for the meter, *P*=.48 for the app, and *P*=.40 for in-app coaching). For participants in the One Drop and activity tracker condition, 95% (40/42) reported the activity tracker was easy to use.

Digital Engagement and Attrition

Among participants that reported on their utilization of One Drop at follow up, 98% (88/90) reported using One Drop's meter, 80% (72/90) reported using the app, and 38% (34/89) reported using in-app coaching half the time or more during the trial to check blood glucose, manage diabetes, or get help with diabetes, respectively. Self-reported use of One Drop during the trial was comparable between groups (*P*=.99 for the meter, *P*=.66 for the app, and *P*=.11 for in-app coaching).

According to objectively collected One Drop user data, 92% (87/95) of all participants used two or more parts of One Drop's 3-part solution during the study. No participant used only one part. One Drop utilization was comparable between groups (χ^2_2=0.45, *P*=.80).

Among participants in the One Drop and activity tracker condition, 100% (42/42) reported using the tracker half the time or more during the study, 90% (38/42) of whom used it always or almost always during that time.

Additionally, 88% (84/95) of all participants completed the study. Study completion did not differ by condition assignment (χ^2_1=0.04, *P*=.83).

Hypoglycemic Events

The Mann-Whitney *U* test suggested a trending difference in the number of hypoglycemic events experienced during the study period between the One Drop and activity tracker group (n=42) and the One Drop only group (n=47). The One Drop and activity tracker group (1.64, SD 3.05) tended to report fewer hypoglycemic events during the study period compared to the One Drop only group (2.98, SD 4.92) (*P*=.08).

ITT Analysis

Descriptive statistics were used to summarize the sample characteristics in the ITT analysis (N=95). The sample was about 40 years old on average, and the majority were female, Caucasian/White race, had at least some college education, an annual household income ≥US $50,000, and were overweight or obese. The sample had received a T1D diagnosis about 20 years ago on average with an average baseline hemoglobin A_{1c} of 8.41% (Table 1).

Twenty pooled ANCOVA models were used to test statistical assumptions on 20 sets of imputed data (N=95; One Drop and activity tracker n=46, One Drop only n=49). Both the homogeneity of regression slopes ($F_{pooled1}$=1.06, P_{pooled}<.42) and homogeneity of variance ($F_{pooled1}$=2.93, P_{pooled}<.11) assumptions were met. Twenty pooled ANCOVA models without the interaction term on 20 sets of imputed data revealed a significant main effect for baseline hemoglobin A_{1c} ($F_{pooled1}$=227.99, P<.001) and a significant main effect for condition assignment ($F_{pooled1}$=10.28, P<.001, ηp^2=.10, observed power=0.87). Follow-up hemoglobin A_{1c} varied by group. After the 3-month study period, participants in the One Drop and activity tracker condition had a significantly lower hemoglobin A_{1c} level (mean 7.9%, SD 0.60%, 95% CI 7.8-8.2) than participants in the One Drop only condition (mean 8.4%, SD 0.62%, 95% CI 8.2-8.5).

Table 1. Characteristics of intent-to-treat participants overall and by group.

Characteristic	All Participants (N=95)	One Drop + tracker (n=46)	One Drop only (n=49)
Demographic characteristics			
Age (years), mean (SD)	40.9 (10.7)	41.1 (9.7)	40.7 (11.6)
Gender, n (%)			
Female	69 (73)	34 (74)	35 (71)
Male	26 (27)	12 (26)	14 (29)
Race/ethnicity, n (%)			
Caucasian/White	84 (88)	43 (94)	41 (84)
Hispanic/Latino	5 (5)	1 (2)	4 (8)
African American/Black	3 (3)	1 (2)	2 (4)
Asian	0 (0)	0 (0)	0 (0)
American Indian/Alaskan Native	2 (2)	0 (0)	2 (4)
Native Hawaiian/Pacific Islander	1 (1)	1 (2)	0 (0)
Education years, mean (SD)	14.6 (2.2)	14.4 (2.2)	14.9 (2.1)
Education level, n (%)			
Below high school	4 (4)	2 (4)	2 (4)
High school graduate/GED[a]	18 (19)	11 (24)	7 (14)
Some college	33 (35)	18 (39)	15 (31)
College graduate	15 (16)	4 (9)	11 (22)
Graduate school	25 (26)	11 (24)	14 (29)
Annual income (USD)[b], n (%)			
<25,000	12 (13)	7 (15)	5 (11)
25,000-50,000	26 (27)	13 (28)	13 (29)
50,000-100,000	35 (37)	21 (46)	14 (31)
>100,000	18 (19)	5 (11)	13 (29)
Health status			
Health insurance, n (%)			
Yes	89 (94)	45 (98)	44 (90)
No	6 (6)	1 (2)	5 (10)
Diabetes duration (years), mean (SD)	20.3 (11.5)	21.6 (12.1)	19.1 (10.9)
BMI, mean (SD)	30.1 (6.6)	30.6 (7.7)	29.6 (5.4)
BMI category, n (%)			
Underweight, BMI<18.5	1 (1)	1 (2)	0 (0)
Normal weight, BMI 18.5-24.9	18 (19)	10 (22)	8 (168)
Overweight, BMI 25-29.9	32 (34)	12 (26)	20 (41)
Obese I, BMI 30-34.9	26 (27)	12 (26)	14 (29)
Obese II, BMI 35-39.9	9 (10)	5 (11)	4 (8)
Morbidly obese, BMI≥40	9 (10)	6 (13)	3 (6)
Hemoglobin A_{1c} (%), mean (SD)	8.4 (1.2)	8.5 (1.9)	8.3 (1.9)

[a]GED: General Educational Development test.

[b]N=91.

PP Analysis

A single ANCOVA model was run on the complete data of participants who used One Drop during the study period and provided a 3-month hemoglobin A_{1c} blood specimen (N=77; One Drop and activity tracker n=36, One Drop only n=41). Once again, the ANCOVA assumptions of homogeneity of regression slopes (F_1=153.3, P<.001) and homogeneity of variance (F_1=5.36, P<.02) were met.

An ANCOVA model without the interaction term revealed a significant main effect for baseline hemoglobin A_{1c} (F_1=153.3, P<.001) and a significant main effect for condition assignment (F_1=5.36, P<.02, ηp^2=.07, observed power=0.63). Follow-up hemoglobin A_{1c} varied by group. Participants who followed protocol during the 3-month study period had a significantly lower hemoglobin A_{1c} level if they were in the One Drop and activity tracker condition (mean 7.9%, SD 0.59%, 95% CI 7.7-8.1) than if they were in the One Drop only condition (mean 8.2%, SD 0.58%, 95% CI 8.0-8.4).

Discussion

Principal Findings

This is the first randomized controlled trial evaluating One Drop with an activity tracker vs One Drop alone on the 3-month hemoglobin A_{1c} of adults with T1D. Participants exposed to One Drop and the tracker for the 3-month study period had a significantly lower 3-month hemoglobin A_{1c} compared to that of participants exposed to One Drop only during the same timeframe. Results were consistent in ITT analyses on imputed data and PP analyses on complete data (ie, among participants using all aspects of One Drop who also provided a follow-up hemoglobin A_{1c} value). Moreover, there was a trend of fewer hypoglycemic events experienced during the study period among participants in the One Drop and activity tracker condition relative to those in the One Drop only condition.

Consistent with behavior change theories and their empirical validations [33,34], self-care improvements may have been a mechanism by which the 3-month hemoglobin A_{1c} was better for participants in the One Drop and activity tracker condition than in the One Drop only condition. People tracking their activity experience small but significant improvements in spontaneous lifestyle activity and weight loss [35]. Additional research is needed to determine if using One Drop and an activity tracker improves physical activity, weight, and, in turn, hemoglobin A_{1c}.

Using hardware and software to track self-care makes people aware of their activity but may not sufficiently engage and activate them [36]. Being female or overweight/obese has been associated with using an activity tracker [37]. The trial's predominantly female (73%) and overweight/obese (80%) sample may have been uniquely engaged and activated by using an activity tracker with One Drop. Therefore, these results may not generalize to adults with T1D who are male or have a "normal" BMI.

Strengths and Limitations

There are study strengths and limitations to note. Prior One Drop studies have been limited by a single group, pre-post design, self-selection, and prior exposure to One Drop as an alternative explanation for the results. This trial's randomized design addresses some of these limitations while highlighting the benefits of activity trackers for people with T1D using One Drop.

This trial was "pragmatic" [38]. That is, the study procedures were conducted remotely in the context of participants' everyday lives. Pragmatic trials maximize the applicability and generalizability of the findings but also make study data open to accuracy concerns. For instance, a person other than the participant might have completed the mail-in hemoglobin A_{1c} test. A real-world trial also introduces more confounding variables. For example, participants may have used other wearable devices or health apps during the study period.

People with diabetes, payers, and manufacturers want to know what solutions to use, purchase, and to whom to market. A more controlled environment reduces alternative explanations for research findings and may have produced more internally valid results, but at the cost of less real-world application. A strength of this trial is that it was remotely conducted and therefore far-reaching. People of different race/ethnicities, social classes, and education levels participated from 43 out of the 50 states.

Participants came from across the United States but may not be representative of people with T1D in the country. Additionally, results may not generalize to specific racial/ethnic minority groups unaccounted for or underrepresented in this study. People with T1D in the United States are disproportionately non-Hispanic White (50%), followed by non-Hispanic Black (30%) and finally Hispanic (18%) and other race/ethnicities (5%) [39]. Our study sample was predominately non-Hispanic White (88%), but with much fewer non-Hispanic Black/Hispanic/Other (12%) racial/ethnic minority participants.

In addition to the study enrolling participants from all over the continental United States, recruitment, data collection, and analyses were also completed in only 6.4 months, saving time, money, and providing just-in-time results to decision makers. It generally takes 17 years to turn 14% of research findings into benefits for patients [40]. This trial strongly challenges how long it takes to conduct a randomized controlled trial and translate results into the real world.

Although the pragmatic and remote study design allowed for recruitment, data collection, and participation to occur in the context of everyday life, making it more convenient than trials requiring study visits at clinical trial sites, remoteness also meant relying on self-reported screening data, and medication and medical history. As a strength, engagement with One Drop (ie, protocol adherence) was objectively determined in real time. In addition, the primary outcome, hemoglobin A_{1c}, was conveniently measured at home, but was processed at a central lab (ie, lab assays are the gold standard).

Implications and Future Research

Better integration is the future of diabetes care. The closed-loop community highlights the desire for integrated tools (ie, people are hacking their insulin pumps and continuous glucose monitors to create their own artificial pancreas [41]). Results from this trial suggest that One Drop and an activity tracker may work better together than alone in helping people with T1D manage their health. To better understand the additive value of One Drop with and without an activity tracker among adults with T1D, future research should include four study arms (complete control, activity tracker only, One Drop only, One Drop and activity tracker) and randomize to condition. The current study was underpowered (N=95) to be a four-arm trial.

Data integration and control are key drivers of the initial and sustained use of consumer health technologies [36,42]. One Drop's app is both integrated with activity trackers and allows users to have control over how they track their data and share it with their One Drop coach or a health care provider. Control, customization, and integration may influence longer-term use and benefit, and should be explored in future research.

Acknowledgments

This work was funded by Informed Data Systems Inc.

Conflicts of Interest

CO and LS were full-time employees of Informed Data Systems Inc (IDS), the manufacturer of One Drop's digital therapeutics solution, during the conduct of this research. AH, MH, BH, and JD are currently full-time employees and have stock in IDS. JR is a member of One Drop's clinical advisory board.

Multimedia Appendix 1
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 2398 KB - mhealth_v8i9e16745_app1.pdf]

References

1. National Diabetes Statistics Report 2017. Centers for Disease Control and Prevention, US Department of Health and Human Services. 2017. URL: https://dev.diabetes.org/sites/default/files/2019-06/cdc-statistics-report-2017.pdf [accessed 2020-09-01]
2. American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2017. Diabetes Care 2018 May;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]
3. Stokes A, Preston SH. Deaths attributable to diabetes in the United States: comparison of data sources and estimation approaches. PLoS One 2017;12(1):e0170219 [FREE Full text] [doi: 10.1371/journal.pone.0170219] [Medline: 28121997]
4. Nathan DM, DCCT/EDIC Research Group. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. Diabetes Care 2014;37(1):9-16 [FREE Full text] [doi: 10.2337/dc13-2112] [Medline: 24356592]
5. DCCT Research Group, Nathan DM, Genuth S, Lachin J, Cleary P, Crofford O, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993 Sep 30;329(14):977-986. [doi: 10.1056/NEJM199309303291401] [Medline: 8366922]
6. Mulvaney SA, Hood KK, Schlundt DG, Osborn CY, Johnson KB, Rothman RL, et al. Development and initial validation of the barriers to diabetes adherence measure for adolescents. Diabetes Res Clin Pract 2011 Oct;94(1):77-83 [FREE Full text] [doi: 10.1016/j.diabres.2011.06.010] [Medline: 21737172]
7. American Diabetes Association. 5. Lifestyle management: standards of medical care in diabetes. Diabetes Care 2019;42(Suppl 1):S46-S60. [doi: 10.2337/dc19-s005]
8. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes. Diabetes Care 2019;42(Suppl 1):S61-S70. [doi: 10.2337/dc19-s006]
9. Al Hayek AA, Robert AA, Babli S, Almonea K, Al Dawish MA. Fear of self-injecting and self-testing and the related risk factors in adolescents with type 1 diabetes: a cross-sectional study. Diabetes Ther 2017 Feb;8(1):75-83 [FREE Full text] [doi: 10.1007/s13300-016-0221-8] [Medline: 27995595]
10. Klonoff DC. Continuous glucose monitoring: roadmap for 21st century diabetes therapy. Diabetes Care 2005 May;28(5):1231-1239. [doi: 10.2337/diacare.28.5.1231] [Medline: 15855600]
11. Weissberg-Benchell J, Antisdel-Lomaglio J, Seshadri R. Insulin pump therapy: a meta-analysis. Diabetes Care 2003 Apr;26(4):1079-1087. [doi: 10.2337/diacare.26.4.1079] [Medline: 12663577]
12. Heintzman ND. A digital ecosystem of diabetes data and technology: services, systems, and tools enabled by wearables, sensors, and apps. J Diabetes Sci Technol 2015 Dec 20;10(1):35-41 [FREE Full text] [doi: 10.1177/1932296815622453] [Medline: 26685994]
13. Barnes TL, Watson DA, Gandrud LM. Higher weekly physical activity in T1D pediatric patients is associated with improved target glucose levels—findings from six months of Fitbit-monitored activity. Diabetes 2018 May;67(Suppl 1):1387-P. [doi: 10.2337/db18-1387-p]

14. Gandrud L, Altan A, Buzinec P, Hemphill J, Chatterton J, Kelley T, et al. Intensive remote monitoring versus conventional care in type 1 diabetes: a randomized controlled trial. Pediatr Diabetes 2018 Feb 21;19(6):1086-1093. [doi: 10.1111/pedi.12654] [Medline: 29464831]
15. Sears LE, Dachis J, Heyman M, Head R, Huddleston B, Osborn CY. People with type 1 diabetes paying for a One Drop coach improve when not at goal, or stay at goal when starting at goal. Ann Behav Med (SBM 39th Annual Meeting Abstract Supplement) 2018;52(Suppl 1):S1-S838. [doi: 10.1093/abm/kay013]
16. Osborn CY, van Ginkel JR, Rodbard D, Heyman M, Marrero DG, Huddleston B, et al. One Drop | Mobile: an evaluation of hemoglobin A1c improvement linked to app engagement. JMIR Diabetes 2017 Aug 24;2(2):e21 [FREE Full text] [doi: 10.2196/diabetes.8039] [Medline: 30291059]
17. Osborn CY. The One Drop diabetes iOS and WatchOS app with in-app coaching from Certified Diabetes Educators improves blood glucose, carbohydrate intake, and physical activity. 2017 Presented at: Stanford MedX; September 15-17, 2017; Palo Alto, CA.
18. Osborn CY, van Ginkel JR, Marrero DG, Rodbard D, Huddleston B, Dachis J. One Drop | mobile on iPhone and Apple watch: an evaluation of HbA1c improvement associated with tracking self-care. JMIR Mhealth Uhealth 2017 Nov 29;5(11):e179 [FREE Full text] [doi: 10.2196/mhealth.8781] [Medline: 29187344]
19. Research2Guidance. Digital Diabetes Care Market 2018-2022: Ready To Take Off. 2018. URL: https://research2guidance.com/product/digital-diabetes-care-market-2018-2022-ready-to-take-off/
20. Klonoff DC, Parkes JL, Kovatchev BP, Kerr D, Bevier WC, Brazg RL, et al. Investigation of the accuracy of 18 marketed blood glucose monitors. Diabetes Care 2018 Aug;41(8):1681-1688. [doi: 10.2337/dc17-1960] [Medline: 29898901]
21. Osborn C, Heyman M, Dachis J. The One Drop Mobile app and Experts program is evidence-based and improves blood glucose. Ann Behav Med (SBM 38th Annual Meeting Abstract Supplement) 2017;51(Suppl 1):S1-S2867.
22. Kim SE, Castro Sweet CM, Gibson E, Madero EN, Rubino B, Morrison J, et al. Evaluation of a digital diabetes prevention program adapted for the Medicaid population: study design and methods for a non-randomized, controlled trial. Contemp Clin Trials Commun 2018 Jun;10:161-168 [FREE Full text] [doi: 10.1016/j.conctc.2018.05.007] [Medline: 30035241]
23. Trief PM, Fisher L, Sandberg J, Cibula DA, Dimmock J, Hessler DM, et al. Health and psychosocial outcomes of a telephonic couples behavior change intervention in patients with poorly controlled type 2 diabetes: a randomized clinical trial. Diabetes Care 2016 Dec;39(12):2165-2173 [FREE Full text] [doi: 10.2337/dc16-0035] [Medline: 27456837]
24. de Boer MR, Waterlander WE, Kuijper LD, Steenhuis IH, Twisk JW. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. Int J Behav Nutr Phys Act 2015 Jan 24;12(1):4 [FREE Full text] [doi: 10.1186/s12966-015-0162-z] [Medline: 25616598]
25. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
26. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. BMC Med Res Methodol 2014 Apr 09;14(1):49 [FREE Full text] [doi: 10.1186/1471-2288-14-49] [Medline: 24712304]
27. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869-c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
28. Ranganathan P, Pramesh C, Aggarwal R. Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. Perspect Clin Res 2016;7(3):144-146 [FREE Full text] [doi: 10.4103/2229-3485.184823] [Medline: 27453832]
29. Detry MA, Lewis RJ. The intention-to-treat principle: how to assess the true effect of choosing a medical treatment. JAMA 2014 Jul 02;312(1):85-86. [doi: 10.1001/jama.2014.7523] [Medline: 25058221]
30. Shrier I, Verhagen E, Stovitz SD. The intention-to-treat analysis is not always the conservative approach. Am J Med 2017 Jul;130(7):867-871. [doi: 10.1016/j.amjmed.2017.03.023] [Medline: 28392186]
31. Brody T. Chapter 8 - Intent-to-Treat Analysis Versus Per Protocol Analysis. In: Clinical Trials, Second Edition. Boston: Academic Press; 2016:253-268.
32. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017 Oct 05;377(14):1391-1398. [doi: 10.1056/nejmsm1605385]
33. Mayberry LS, Osborn CY. Empirical validation of the information-motivation-behavioral skills model of diabetes medication adherence: a framework for intervention. Diabetes Care 2014;37(5):1246-1253 [FREE Full text] [doi: 10.2337/dc13-1828] [Medline: 24598245]
34. Osborn CY, Paasche-Orlow MK, Bailey SC, Wolf MS. The mechanisms linking health literacy to behavior and health status. Am J Health Behav 2011;35(1):118-128 [FREE Full text] [doi: 10.5993/ajhb.35.1.11] [Medline: 20950164]
35. Goode AP, Hall KS, Batch BC, Huffman KM, Hastings SN, Allen KD, et al. The impact of interventions that integrate accelerometers on physical activity and weight loss: a systematic review. Ann Behav Med 2017 Feb;51(1):79-93 [FREE Full text] [doi: 10.1007/s12160-016-9829-1] [Medline: 27565168]
36. Miyamoto SW, Henderson S, Young HM, Pande A, Han JJ. Tracking health data is not enough: a qualitative exploration of the role of healthcare partnerships and mHealth technology to promote physical activity and to sustain behavior change. JMIR Mhealth Uhealth 2016 Jan 20;4(1):e5 [FREE Full text] [doi: 10.2196/mhealth.4814] [Medline: 26792225]

37. Macridis S, Johnston N, Johnson S, Vallance JK. Consumer physical activity tracking device ownership and use among a population-based sample of adults. PLoS One 2018;13(1):e0189298 [FREE Full text] [doi: 10.1371/journal.pone.0189298] [Medline: 29293532]
38. Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? BMC Med 2018 Apr 03;16(1):49 [FREE Full text] [doi: 10.1186/s12916-018-1038-2] [Medline: 29615035]
39. Menke A, Orchard TJ, Imperatore G, Bullard KM, Mayer-Davis E, Cowie CC. The prevalence of type 1 diabetes in the United States. Epidemiology 2013;24(5):773-774. [doi: 10.1097/ede.0b013e31829ef01a]
40. Balas E, Weingarten S, Garb CT, Blumenthal D, Boren S, Brown GD. Improving preventive care by prompting physicians. Arch Intern Med 2000 Feb 14;160(3):301-308. [doi: 10.1001/archinte.160.3.301] [Medline: 10668831]
41. Weaver KW, Hirsch IB. The hybrid closed-loop system: evolution and practical applications. Diabetes Technol Ther 2018 Jun;20(S2):S216-S223. [doi: 10.1089/dia.2018.0091] [Medline: 29873517]
42. Fukuoka Y, Lindgren T, Jong S. Qualitative exploration of the acceptability of a mobile phone and pedometer-based physical activity program in a diverse sample of sedentary women. Public Health Nurs 2012;29(3):232-240. [doi: 10.1111/j.1525-1446.2011.00997.x]

Abbreviations

ANCOVA: analysis of covariance
CDE: Certified Diabetes Educator
CONSORT: Consolidated Standards of Reporting Trials
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
IRB: Institutional Review Board
ITT: intention to treat
PP: per protocol
T1D: type 1 diabetes

Edited by G Eysenbach; submitted 20.10.19; peer-reviewed by D Ahn, L Quinlan, C Mamillapalli; comments to author 19.11.19; revised version received 24.01.20; accepted 13.06.20; published 17.09.20.

Please cite as:
Osborn CY, Hirsch A, Sears LE, Heyman M, Raymond J, Huddleston B, Dachis J
One Drop App With an Activity Tracker for Adults With Type 1 Diabetes: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(9):e16745
URL: http://mhealth.jmir.org/2020/9/e16745/
doi:10.2196/16745
PMID:32540842

Original Paper

Effect of Voluntary Participation on Mobile Health Care in Diabetes Management: Randomized Controlled Open-Label Trial

Da Young Lee[1,2*], MD; Seung-Hyun Yoo[2,3*], MD; Kyong Pil Min[4], MS; Cheol-Young Park[1], PhD

[1]Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[2]Korea University College of Medicine, Seoul, Republic of Korea

[3]National Health Insurance Service, Wonju-Si, Republic of Korea

[4]Huraypositive Inc, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Cheol-Young Park, PhD
Kangbuk Samsung Hospital
Sungkyunkwan University School of Medicine
29 Saemunan-ro, Jongno-gu
Seoul, 03181
Republic of Korea
Phone: 82 220011869
Email: cydoctor@chol.com

Abstract

Background: The role of mobile health care (mHealth) in glycemic control has been investigated, but its impact on self-management skills and its psychological aspects have not been studied.

Objective: We evaluated the efficacy of mHealth-based diabetes self-management education and the effect of voluntary participation on its effects.

Methods: This study was a randomized controlled open-label trial conducted for 6 months at Kangbuk Samsung Hospital. Participants in the control group (n=31) maintained their previous diabetes management strategies. Participants in the intervention group (n=41) additionally received mHealth-based diabetes self-management education through a mobile app and regular individualized feedback from health care professionals. The primary outcome was change in glycated hemoglobin (HbA_{1c}) level over 6 months between the 2 groups (intervention versus control) and within each group (at 6 months versus baseline). The secondary outcomes were changes in body mass index, blood pressure, lipid profile, and questionnaire scores (the Korean version of the Summary of Diabetes Self-Care Activities Questionnaire, an Audit of Diabetes Dependent Quality of Life, the Appraisal of Diabetes Scale, and Problem Areas in Diabetes) over 6 months between groups and within each group.

Results: A total of 66 participants completed this study. HbA_{1c} (P=.04), total cholesterol level (P=.04), and Problem Areas in Diabetes scores (P=.02) significantly decreased; total diet (P=.03) and self-monitoring of blood glucose level scores (P=.01), based on the Summary of Diabetes Self-Care Activities Questionnaire, markedly increased within the intervention group. These significant changes were observed in self-motivated participants who were recruited voluntarily via advertisements.

Conclusions: mHealth-based diabetes self-management education was effective at improving glycemic control and diabetes self-management skills and lowering diabetes-related distress in voluntary participants.

Trial Registration: ClinicalTrials.gov NCT03468283; http://clinicaltrials.gov/ct2/show/NCT03468283

(JMIR Mhealth Uhealth 2020;8(9):e19153) doi:10.2196/19153

KEYWORDS

diabetes mellitus; health services research; mobile applications; diabetes; mHealth; app; lifestyle; self-management; volunteer; participation

Introduction

Lifestyle management, including diabetes self-management education, nutrition therapy, and physical activity, is a fundamental aspect of the successful management of diabetes [1,2]. Previous studies [1] have demonstrated that diabetes self-management education improves self-management, diabetes knowledge, satisfaction and quality of life, and glycemic control, and also reduces health care costs. However, only a small number of individuals eligible for diabetes self-management education receive it [3], and poor adherence has been reported [4]. Given that barriers to adherence might include patient factors, a patient-centered approach that would improve self-motivation could be useful.

Mobile health care (mHealth)–based diabetes self-management education has been regarded as an innovative option for diabetes self-management education, in that it may help to overcome time and location barriers and provide real-time individualized medical treatments [5-9]. The effectiveness of mHealth in diabetes management has been demonstrated in a few studies [6,10-13]. However, these previous studies [6,10-13] have mainly focused on the role of self-management education in glycemic control; the impact of mHealth on self-management skills and its psychological aspects have not been studied.

Therefore, we conducted this study to evaluate whether implementing mHealth-based diabetes self-management education could improve diabetes self-management and glycemic control and enhance patient quality of life. In addition, we aimed to investigate the importance of voluntary engagement in mHealth-based diabetes self-management education.

Methods

Study Participants

Patients with type 2 diabetes from an outpatient clinic at Kangbuk Samsung Hospital were invited to participate via advertisement in Kangbuk Samsung Hospital or via recommendation by their physician between June 2012 and July 2012 (Multimedia Appendix 1). Inclusion criteria were as follows: age ≥19 years; Android smartphone users; no changes in diabetes medication for at least 6 months; and HbA_{1c} levels ≥6.5% within the last 3 months. We recruited a total of 140 individuals (70 individuals for each group), such that the sample size afforded 80% power at a significance level of α=.05 with a 10% dropout rate, considering a mean difference in HbA_{1c} level of –0.50 based on prior research [5] investigating similar interventions. However, because many participants were excluded because they had an iPhone rather than an Android-based phone or had changed medications within 6 months, only 73 individuals were initially found to be eligible. Among these 73 individuals, 1 participant was excluded. Exclusion criteria were having a serious concomitant disease other than diabetes; a history of malignancy, myocardial infarction, cerebral infarction, or organ transplantation; being pregnant or planning for pregnancy within 6 months; planning to participate in other clinical studies; or illiteracy. Finally, 72 individuals were enrolled in this study.

All participants provided written informed consent before study procedures began. This study was reviewed and approved by the institutional review board of Kangbuk Samsung Hospital (KBS12089) and was carried out in accordance with the Helsinki Declaration. The trial was registered on February 26, 2018 with clinicaltrials.gov (NCT03468283). The CONSORT checklist [14] is available in Multimedia Appendix 2.

Study Design and Details of the Intervention

This study was a randomized controlled open-label trial conducted from June 2012 to March 2013 at Kangbuk Samsung Hospital, Sungkyunkwan University, Seoul, Republic of Korea. Participants were randomly assigned to 1 of 2 groups: the intervention group or the control group. We used a computer-generated list of random numbers produced by a statistician with no clinical involvement in the trial to allocate the participants. Participants were allocated in the order in which they were registered.

Participants in the intervention group received mHealth-based diabetes self-management education, which consisted of the mobile app Healthynote (for Android; CVnet Co) and regular individualized feedback messages from health care professionals regarding their diabetes management. The intervention group uploaded the Healthynote app onto their own smartphone, were educated on how to use this service, and were provided a near field communication–enabled glucometer (CareSens N NFC; i-SENS Inc). They entered their medical information, such as self-monitored blood glucose level, dietary record, exercise, blood pressure, medication record, and body weight into the app and could check their data at any time by logging into the mobile app and could share it through social network services. The data entered in the app were automatically transmitted by wireless network to the server and stored on a secure website accessible only to providers. During the 6-month study period, participants in the intervention group received regular mobile messages from health care professionals, consisting of 2 endocrinologists and a nurse, via the mobile app once or twice a week. Messages contained general information about diabetes (eg, medications, diet, exercise, treatment goals), diabetes self-care behaviors based on recommendations from the American Diabetes Association [15] and the Korean Diabetes Association [16], encouragement, reminders if the patient had not used the app recently, and individualized advice based on their entered data. Participants could communicate bidirectionally with providers via app messages and call server administrators for technical problems. In the secure website for health care professionals, participant-generated health data were tracked and analyzed for trends and patterns. Using a function for sorting and filtering participants according to transmitted data, usage, and scheduled period, providers sent messages.

Participants in the control group maintained their previous diabetes management at Kangbuk Samsung Hospital. Providers were not involved with patient prescriptions. Among 7 endocrinologists at this center, 2 participated in this study as the principal investigator and co-investigator. Participants were recruited regardless of their endocrinologist. Providers telephoned participants in both groups to discuss research

progress. Patients in the study did not have to make additional visits to the hospital beyond regular visits.

To evaluate the impact of self-motivation among our study participants, we divided the intervention group into 2 subgroups for analysis: a group of participants that voluntarily participated in the mobile diabetes self-management education program and joined after seeing an advertisement posted in the hospital (ie, the self-referred group, n=20) and a group who joined based on recommendation by their primary care physician (ie, the physician-referred group, n=19).

Anthropometric and Laboratory Measurements

At baseline, all participants completed a self-administered questionnaire regarding demographic characteristics, social history, and other medical conditions. BMI was calculated as weight in kilograms divided by the square of height in meters. Blood pressure was measured twice using a standardized sphygmomanometer after 5 minutes of rest, and the lower values were used.

Venous blood samples were collected in the morning (between 8 AM and 9 AM) after an overnight fast of more than 8 hours. Concentrations of plasma glucose (fasting plasma glucose) were determined using the hexokinase method; glycated hemoglobin (HbA_{1c}) level was determined using turbidimetric inhibition immunoassay; and lipid levels were determined using an enzymatic method and a homogeneous enzymatic calorimetric test. Biochemical values were measured using the Cobas Modular 6000 analyzer series (Roche Diagnostics). Methodology was aligned with Diabetes Control and Complications Trial and National Glycohemoglobin Standardization Program standards [17]. The coefficients of variation were 1.94% and 2.2% for HbA_{1c}, 9.57% and 5.39% for cholesterol, and 7.54% and 4.14% for triglyceride level, for normal and abnormal values, respectively.

These variables were tested both at baseline and at 6 months. In addition, blood pressure, weight, lipid profile, fasting plasma glucose level, and HbA_{1c} level were also measured at 3 months.

Questionnaires

To assess diabetes self-management, the impact of diabetes on quality of life, diabetes awareness, emotional stress derived from diabetes, and treatment satisfaction, participants answered 5 self-administered questionnaires both at baseline and at 6 months. The Korean version of the Summary of Diabetes Self-Care Activities Questionnaire (SDSCA) [18,19], the Audit of Diabetes Dependent Quality of Life (ADDQOL) [20,21], the Korean version of the Appraisal of Diabetes Scale (ADS) [22,23], the Problem Areas in Diabetes (PAID) questionnaire [24], and Diabetes Treatment Satisfaction Questionnaire status version (DTSQs) and change version (DTSQc) [25,26] were used.

Statistical Analysis

We conducted a per protocol analysis. The primary outcomes were (1) difference (between the intervention and control groups) in HbA_{1c} levels at 6 months and (2) change (within each group) in HbA_{1c} levels from baseline at 6 months.

Two-tailed independent *t* tests, paired *t* tests, and chi-square tests were used to examine differences between the groups.

The comparisons of secondary outcomes—BMI, systolic blood pressure, diastolic blood pressure, lipid levels, and SDSCA, ADDQOL, ADS, and PAID scores between the 2 groups and within each group—were performed using Wilcoxon signed-rank tests and paired *t* tests. We compared the DTSQ scores of the 2 groups without analyzing longitudinal differences.

Additionally, we repeated the above-mentioned analysis after dividing the participants into control, voluntary, and physician-referred groups.

We calculated ΔHbA_{1c} by subtracting baseline HbA_{1c} level from the HbA_{1c} level measured at 6 months (lower ΔHbA_{1c} indicated better glycemic control). We conducted multivariate linear regression analysis adjusted for age, sex, baseline BMI, systolic blood pressure, HbA_{1c}, high-density lipoprotein cholesterol level, and low-density lipoprotein cholesterol level to assess the relationship between ΔHbA_{1c} and the total frequency of self-monitoring of blood glucose level records, exercise records, diet records, medication records, and message reading rate for 6 months.

SPSS statistical software (version 22.0; IBM Corp) was used. A value of $P<.05$ was considered statistically significant. Bonferroni correction (for multiple comparisons) was used.

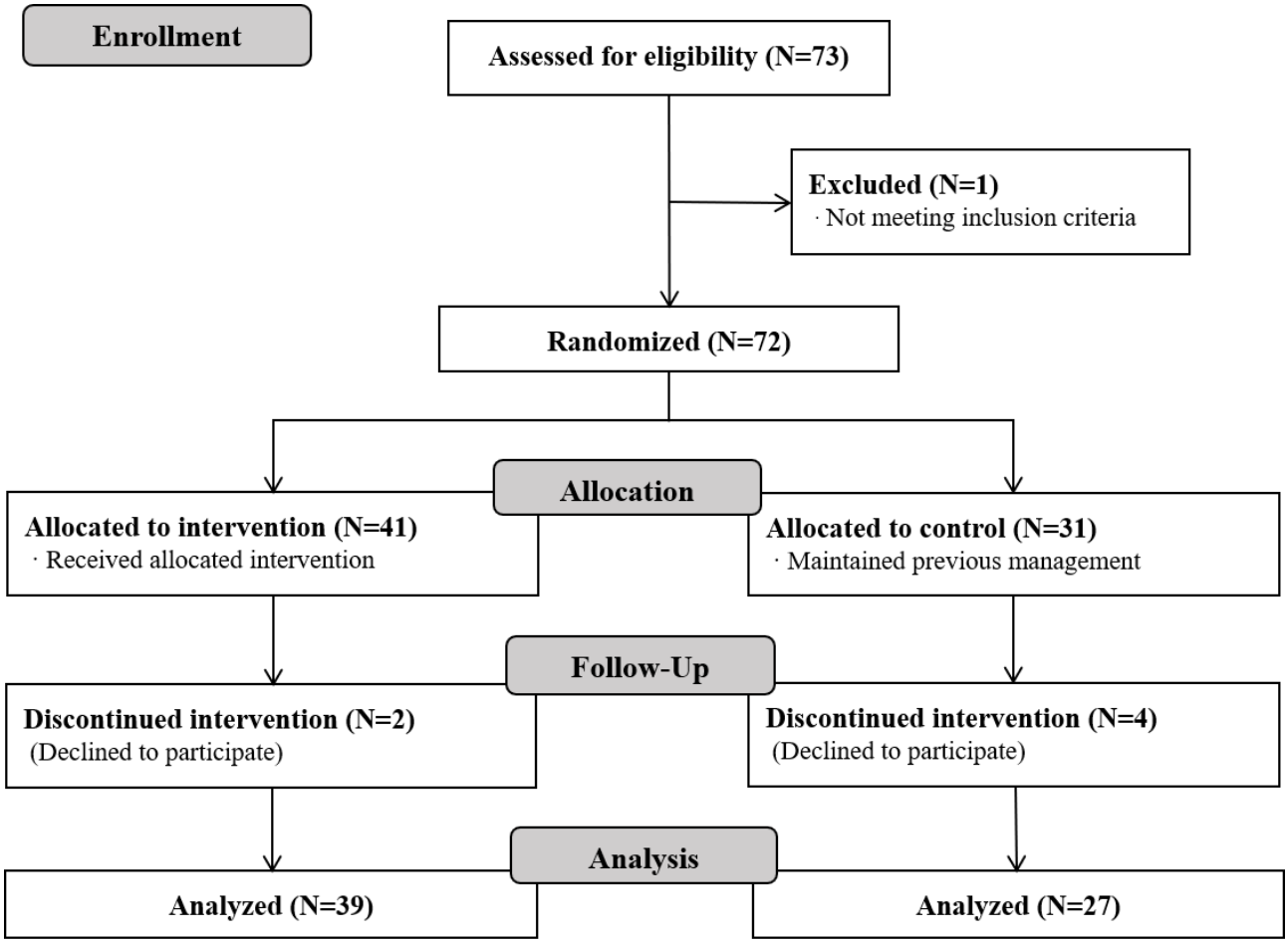
Results

Study Execution and Baseline Characteristics of Participants

Among 72 participants, 66 completed the study (Figure 1); 2 participants from the intervention group and 4 participants from the control group were lost to follow-up.

There were no differences in metabolic parameters at baseline (Multimedia Appendix 3). Diabetes self-management status, especially diet and self-monitored blood glucose level, was superior in the control group.

Figure 1. CONSORT flow diagram.



Changes in Metabolic Parameters and Questionnaire Scores

Although there was no significant difference between the control and intervention groups at 6 months (HbA_{1c}: *P*=.05; total cholesterol *P*=.24), the intervention group showed statistical improvement in HbA_{1c} (*P*=.04) and total cholesterol levels (*P*=.04) after 6 months (Multimedia Appendix 4). SDSCA scores, especially total diet (*P*=.03) and self-monitoring of blood glucose level (*P*=.01), which were markedly increased compared with baseline, and DTSQ (change) scores at 6 months, which were significantly higher than those of the control group (*P*=.02). The PAID score significantly decreased during the study period in the intervention group, while there was no change in the control group.

As shown in Tables 1 and 2, the significant findings observed in the intervention group were concentrated in the self-referred group. Compared to baseline, HbA_{1c}, total cholesterol levels, low-density lipoprotein cholesterol levels, total diet, and self-monitoring of blood glucose level on the SDSCA were only improved in the self-referred group. Compared to the physician-referred group, the self-referred group also had lower total cholesterol levels, triglyceride levels, exercise, and smoking scores on the SDSCA at 6 months.

Table 1. Changes in biochemical parameters over 6 months.

Parameter	Control (n=27)	Intervention			*P* value[a]		
		All (n=39)	Physician-referred (n=19)	Self-referred (n=20)	Physician-referred vs. control	Self-referred vs. control	Physician- vs. self-referred
HbA_{1c}[b] (%)							
Baseline	7.5 (1.1)	7.4 (0.8)	7.3 (0.7)	7.4 (0.9)	.65	.92	.59
3 months	7.3 (0.7)	7.1 (0.9)	7.0 (0.7)	7.2 (1.0)	.54	.94	.78
6 months	7.6 (0.9)	7.1 (0.8)	7.2 (0.7)	7.1 (1.0)	.13	.09	.73
P value[c,d]	.87	.04	.46	.04			
BMI (kg/m^2)							
Baseline	25.5 (3.0)	26.8 (4.2)	27.3 (4)	26.3 (3.7)	.24	.76	.44
6 months	25.7 (3.1)	27.0 (4.1)	27.7 (4.7)	26.3 (3.4)	.09	.50	.31
P value[c]	.08	.19	.06	.89			
Systolic blood pressure (mmHg)							
Baseline	120.5 (12.0)	121.1 (14.1)	123.6 (13.5)	118.7 (14.5)	.71	.89	.28
6 months	120.2 (11.0)	124.4 (12.0)	128.3 (11.6)	120.7 (11.4)	.02	.89	.05
P value[c]	.91	.07	.10	.40			
Total cholesterol (mg/dL)							
Baseline	150.2 (26.5)	148.7 (41.6)	156.2 (45.3)	141.5 (37.7)	.85	.70	.29
6 months	149.5 (30.2)	139.8 (33.2)	153.4 (38.3)	128.4 (23.5)	.71	.01	.02
P value[c]	.88	.04	.45	.04			
Triglyceride level (mg/dL)							
Baseline	148.2 (67.5)	151.0 (113.5)	182.6 (143.6)	121.1 (65.9)	.46	.61	.11
6 months	216.9 (206.6)	177.9 (116.8)	234.8 (146.3)	129.6 (49.9)	.76	.07	.01
P value[c]	.10	.20	.18	.91			
LDL C[e] (mg/dL)							
Baseline	76.4 (18.0)	77.5 (25.6)	76.0 (23.1)	79.0 (28.3)	.99	.93	.73
6 months	76.1 (18.2)	70.9 (20.7)	74.7 (24.1)	67.6 (17.2)	.83	.11	.30
P value[c]	.91	.08	.77	.02			

[a]An independent *t* test was used.

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

[c]A paired *t* test or Wilcoxon signed-rank test was used.

[d]*P* value is for comparison between baseline and 6 months.

[e]LDL C: low-density lipoprotein cholesterol.

Table 2. Changes in the 5 questionnaires over 6 months.

Parameter	Control (n=27)	Intervention			*P* value[a]		
		All (n=39)	Physician-referred (n=19)	Self-referred (n=20)	Physician-referred vs. control	Self-referred vs. control	Physician- vs. self-referred
SDSCA[b]							
Total diet							
Baseline	14.3 (4.4)	10.8 (3.9)	11.6 (3.0)	10.0 (4.4)	.11	<.001	.22
6 months	14.7 (5.2)	12.6 (5.3)	11.3 (3.3)	13.8 (6.5)	.01	.59	.15
P value[c]	.73	.03	.87	.02			
Exercise							
Baseline	7.2 (3.8)	6.0 (3.5)	5.4 (3.3)	6.6 (3.7)	.27	.84	.33
6 months	6.2 (4.4)	5.2 (3.7)	3.8 (2.7)	6.5 (4.1)	.03	.84	.02
P value[c]	.45	.16	.21	.44			
Self-monitoring of blood glucose							
Baseline	7.1 (4.7)	4.1 (4.9)	3.8 (5.4)	4.4 (4.5)	.10	.19	.42
6 months	6.7 (4.5)	6.6 (4.7)	5.1 (4.2)	8.0 (4.8)	.23	.36	.06
P value[c]	.35	.01	.24	.02			
Foot							
Baseline	4.1 (4.0)	3.8 (4.3)	3.2 (3.8)	4.3 (4.7)	.77	.98	.44
6 months	4.1 (4.0)	3.8 (4.3)	3.2 (3.8)	4.3 (4.7)	.46	.86	.43
P value[b]	>.999	.10	>.999	>.999			
Smoking							
Baseline	1.0 (2.2)	2.1 (3.2)	2.8 (3.5)	1.5 (2.9)	.10	.83	.20
6 months	1.7 (3.0)	2.1 (3.2)	3.2 (3.5)	1.1 (2.6)	.16	.41	.04
P value[c]	.10	.46	.18	.16			
ADDQOL[d]							
Baseline	–3.3 (2.1)	–2.5 (1.5)	–2.7 (1.6)	–2.4 (1.6)	.54	.26	.59
6 months	–3.2 (1.7)	–2.8 (1.4)	–2.6 (1.4)	–2.9 (1.5)	.31	.59	.62
P value[c]	.58	.59	.67	.53			
ADS[e] total							
Baseline	15.5 (2.7)	15.8 (3.0)	15.7 (2.0)	15.9 (3.7)	.99	.91	.82
6 months	16.1 (2.4)	15.5 (3.2)	15.2 (2.7)	15.8 (3.6)	.24	.71	.58
P value[c]	.31	.48	.32	.86			
PAID[f]							
Baseline	50.6 (9.8)	49.8 (15.2)	49.9 (13.5)	49.8 (16.9)	.98	.98	.99
6 months	47.5 (12.6)	44.2 (15.2)	42.4 (14.8)	45.9 (15.7)	.22	.70	.49
P value[c]	.30	.02	.09	.18			
DTSQ[g]							
Baseline (DTSQs)	25.2 (4.7)	24.7 (6.1)	24.2 (6.5)	25.3 (5.8)	.80	.99	.59
6 months (DTSQc)	9.0 (4.6)	12.0 (4.7)	11.4 (4.6)	12.6 (4.9)	.10	.02	.47

[a]An independent *t* test was used.

[b]SDSCA: Summary of Diabetes Self-Care Activities.

[c]A paired *t* test or Wilcoxon signed-rank test was used.

[d]ADDQOL: Audit of Diabetes Dependent Quality of Life.

[e]ADS: Appraisal of Diabetes Scale (Korean version).

[f]PAID: Problem Areas In Diabetes.

[g]DTSQc: Diabetes Treatment Satisfaction Questionnaire.

Association Between Record Usage and ΔHbA_{1c}

During the 6-month study period, a mean of 8.0 (SD 7.1) records per week of self-monitored blood glucose level, a mean of 0.9 (SD 1.6) records per week of exercise, and a mean of 1.5 (SD 1.6) records per week of diet were uploaded, and the message reading rate was a mean 54.0% (SD 33.4%). Participants in the self-referred group used the mobile app more than those in the physician-referred group (Table 3). In multivariate linear regression analysis, the correlations of self-monitored blood glucose coefficient, exercise record, diet record, and message reading rate with ΔHbA_{1c} were –0.002 (*P*<.001), –0.008 (*P*=.08), –0.005 (*P*=.07), and –0.005 (*P*=.10), respectively.

Table 3. Mobile app usage and correlation with ΔHbA_{1c} in the intervention group.

Factors	Physician-referred (n=19)	Self-referred (n=20)	*P* value[a]	Standardized coefficient (β)	*P* value[b]
Self-monitoring of blood glucose level record (per week)	5.3 (5.0)	10.6 (8.0)	.02	–0.442	.01
Exercise record (per week)	0.4 (0.9)	1.4 (1.9)	.04	–0.308	.08
Diet record (per week)	1.3 (1.7)	1.7 (1.5)	.43	–0.312	.07
Message reading rate	46.9 (33.2)	61.1 (33.0)	.19	–0.273	.10

[a]An independent *t* test was used.

[b]Linear regression with adjustments for age, sex, and baseline BMI, systolic blood pressure, HbA_{1c} level, high-density lipoprotein cholesterol level, and low-density lipoprotein cholesterol level.

Adverse Events

There were 4 adverse events (2 in the intervention group and 2 in the control group) during the study period: an episode of depression, hospital admission due to aggravation of glycemic control, herpes zoster, and a traffic accident. However, these 4 patients completed the study.

Discussion

This prospective study showed that mHealth-based diabetes self-management education helps to improve glycemic status and diabetes self-management skills, reduces diabetes-related distress, and provides higher user satisfaction, especially in voluntary participants.

Although over 1100 diabetes-related smartphone apps are currently available [27], and prior studies [6,10-13,28,29] have emphasized their impact on HbA_{1c}, there are limited data available on improving self-care skills or other clinical endpoints in patients with type 2 diabetes [30].

In other chronic diseases such as heart failure, chronic lung disease, or cardiovascular disease, mHealth technologies have demonstrated potential for facilitating adherence to chronic disease management [31], as well as improving lipid levels and weight control [32,33].

While mHealth-based diabetes self-management education provides helpful reminders and behavioral reinforcements for adherence to complex care regimens such as medications, blood glucose level checking, and secondary prevention testing [5], weekly health counseling via app messages provides regular contact with health professionals and could have strengthened self-management and behavior changes [34], in addition to offering emotional support and detailed reflection. The characteristics of this study were in line with the 4 key elements of technology-based diabetes management interventions for HbA_{1c} decrement that were suggested by Greenwood et al [35]: (1) 2-way communication, (2) analysis of patient-generated health data, (3) tailored education, and (4) individualized feedback.

Most clinically meaningful changes observed in the intervention group were seen in the voluntary (active engagement) group (Tables 1 and 2). The proportion of primary care physicians who participated in this study as investigators in the self-referred group (2/20, 10%) was significantly lower (*P*<.001 from chi-square analysis) than that of the physician-referred group (13/19, 68.4%). This reflects a significantly higher level of self-motivation for diabetes management among participants within the self-referred group. These findings are in line with previous evidence that initial active engagement in self-monitoring with a telemonitoring device contributed to excellent glycemic improvement [36].

After a slight decrement in HbA_{1c} levels over the first 3 months (Table 1), this improvement was maintained only in the self-referred group at 6 months, resulting in a significant decrement versus baseline (*P*=.04); this phenomenon is referred to as the Hawthorne effect [37]. Significance of the self-referred group (*P*=.04) was related to the impact of mHealth-based diabetes self-management education. This characteristic is important to understanding who uses apps and what keeps them engaged [38] and can be used to select participants for

large-population mHealth services in the future. Advanced model cycling in intensive mHealth-based diabetes self-management education and a self-application period after selecting highly active individuals will be commercially applicable to managing more participants in the future. The degree of improvement in HbA_{1c} was less than that seen for other mHealth models for diabetes [5,6,10-13]. However, as a higher baseline HbA_{1c} level is correlated with a greater decrease in HbA_{1c} in response to various antidiabetes management methods [39], a slight improvement in HbA_{1c} level was predicted in this study, which had a mean baseline HbA_{1c} level of 7.4% (SD 0.8%). In prior studies with a baseline HbA_{1c} level <8.0%, the decrement in HbA_{1c} was approximately 0.4% [6,10].

Frequent app use, especially frequent self-monitored blood glucose level, showed a notable positive impact in our study, as reported in previous studies [40]. Through the ease and accuracy of blood glucose level tracking for patients via a near field communication–enabled glucometer, participants were encouraged to perform more self-monitored blood glucose level. There was also a possible correlation between the number of exercise records and diet records with ΔHba_{1c}, although nonsignificance was indicated (P=.08 and P=.07, respectively; Table 3).

Despite the previously mentioned strengths of this study, several limitations should be considered. First, the number of participants was small, and the researchers were not blinded, given that this study was an open-label study. Second, unlike the selection of the intervention group, the stratification of physician-referred and self-referred groups could not be randomized. The physician-referred and self-referred groups might differ in terms of their familiarity with the use of digital equipment, as participant digital proficiency and comprehension of the mHealth-based diabetes self-management education were not considered at baseline or during the follow-up period. Third, we conducted per-protocol analysis, rather than intention-to-treat analysis. There were no data available after the baseline exam for the 2 participants who declined to participate in intervention group. Because we thought that mHealth-based diabetes self-management education was applied in those participants, we decided to exclude them from analysis. Fourth, this trial was registered at clinicaltrials.gov after the end of the study. Finally, although the real comparator between unmotivated and motivated groups would be among people who declined to participate in this study, we have no way of obtaining information about those patients.

In conclusion, this prospective study demonstrated that an mHealth-based diabetes self-management education that supports diabetes self-management through health counseling messages leads to improved glycemic status and diabetes self-management skills, reduced diabetes-related distress, and high user satisfaction. These positive impacts were predominantly observed in participants who engaged in this mobile app in a voluntary manner and in those who utilized the app more frequently.

Acknowledgments

CVnet Co funded the study, and i-SENS Inc provided glucometers. However, these funders had no role in the design, conduct, or analysis of the trial.

Authors' Contributions

S-HY and C-YP contributed to study design and data analysis. KPM provided technical support for the study and collected data. DYL and S-HY collected and organized data, and wrote, reviewed, and edited the manuscript. C-YP participated in the analytic discussion of the results, supervised, and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

KPM was an employee of CVnet Co during the study and is now an employee of Huraypositive Inc. DYL, S-HY, and C-YP have no conflicts of interest to declare.

Multimedia Appendix 1
Study design.
[PNG File , 1374 KB - mhealth_v8i9e19153_app1.png]

Multimedia Appendix 2
CONSORT checklist.
[PDF File (Adobe PDF File), 1535 KB - mhealth_v8i9e19153_app2.pdf]

Multimedia Appendix 3
Baseline characteristics and comparisons between the three groups.
[DOCX File , 22 KB - mhealth_v8i9e19153_app3.docx]

Multimedia Appendix 4

Changes in biochemical parameters and six questionnaires in two groups over six months.
[DOCX File , 21 KB - mhealth_v8i9e19153_app4.docx]

References

1. American Diabetes Association. 4. Lifestyle Management: Standards of Medical Care in Diabetes-2018. Diabetes Care 2018 Jan;41(Suppl 1):S38-S50. [doi: 10.2337/dc18-S004] [Medline: 29222375]
2. Kim H, Kim H, Yang H, Lee EY, Jeong YJ, Kim TM, et al. Physician-Directed Diabetes Education without a Medication Change and Associated Patient Outcomes. Diabetes Metab J 2017;41(3):187. [doi: 10.4093/dmj.2017.41.3.187]
3. Li R, Shrestha SS, Lipman R, Burrows NR, Kolb LE, Rutledge S, et al. Diabetes self-management education and training among privately insured persons with newly diagnosed diabetes–United States, 2011-2012. MMWR Morb Mortal Wkly Rep 2014 Nov 21;63(46):1045-1049 [FREE Full text] [Medline: 25412060]
4. Bailey CJ, Kodack M. Patient adherence to medication requirements for therapy of type 2 diabetes. Int J Clin Pract 2011 Mar;65(3):314-322. [doi: 10.1111/j.1742-1241.2010.02544.x] [Medline: 21314869]
5. Garabedian LF, Ross-Degnan D, Wharam JF. Mobile Phone and Smartphone Technologies for Diabetes Care and Self-Management. Curr Diab Rep 2015 Dec;15(12):109. [doi: 10.1007/s11892-015-0680-8] [Medline: 26458380]
6. Lim S, Kang SM, Shin H, Lee HJ, Won YJ, Yu SH, et al. Improved glycemic control without hypoglycemia in elderly diabetic patients using the ubiquitous healthcare service, a new medical information system. Diabetes Care 2011 Feb;34(2):308-313 [FREE Full text] [doi: 10.2337/dc10-1447] [Medline: 21270188]
7. Lee DY, Park J, Choi D, Ahn H, Park S, Park C. The effectiveness, reproducibility, and durability of tailored mobile coaching on diabetes management in policyholders: A randomized, controlled, open-label study. Sci Rep 2018 Feb 26;8(1):3642 [FREE Full text] [doi: 10.1038/s41598-018-22034-0] [Medline: 29483559]
8. Lee M, Lee K, Yoo S, Park C. Impact of initial active engagement in self-monitoring with a telemonitoring device on glycemic control among patients with type 2 diabetes. Sci Rep 2017 Jun 20;7(1):3866 [FREE Full text] [doi: 10.1038/s41598-017-03842-2] [Medline: 28634381]
9. Sung J, Lee DY, Min KP, Park C. Peripartum management of gestational diabetes using a digital health care service: a pilot, randomized controlled study. Clin Ther 2019 Nov;41(11):2426-2434. [doi: 10.1016/j.clinthera.2019.09.005] [Medline: 31587813]
10. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
11. 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;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
12. Charpentier G, Benhamou P, Dardari D, Clergeot A, Franc S, Schaepelynck-Belicar P, et al. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). Diabetes Care 2011 Mar;34(3):533-539 [FREE Full text] [doi: 10.2337/dc10-1259] [Medline: 21266648]
13. 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]
14. Eysenbach G, CONSORT- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
15. NGSP. Harmonizing hemoglobin A1c testing. Internet. 2014. URL: http://www.ngsp.org/docs/Protocol.pdf [accessed 2017-05-24] [WebCite Cache ID 6qghVrNFE]
16. Ko SH, Hur KY, Rhee SY, Kim NH, Moon MK, Park SO, Committee of Clinical Practice Guideline of Korean Diabetes Association. Antihyperglycemic Agent Therapy for Adult Patients with Type 2 Diabetes Mellitus 2017: A Position Statement of the Korean Diabetes Association. Diabetes Metab J 2017 Oct;41(5):337-348 [FREE Full text] [doi: 10.4093/dmj.2017.41.5.337] [Medline: 29086531]
17. NGSP. List of NGSP certified methods. Internet. 2016. URL: http://www.ngsp.org/docs/methods.pdf [accessed 2018-07-23]
18. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950 [FREE Full text] [Medline: 10895844]
19. Choi EJ, Nam M, Kim SH, Park CG, Toobert DJ, Yoo JS, et al. Psychometric properties of a Korean version of the summary of diabetes self-care activities measure. Int J Nurs Stud 2011 Mar;48(3):333-337. [doi: 10.1016/j.ijnurstu.2010.08.007] [Medline: 20950807]
20. Wee H, Tan C, Goh S, Li S. Usefulness of the Audit of Diabetes-Dependent Quality-of-Life (ADDQoL) questionnaire in patients with diabetes in a multi-ethnic Asian country. Pharmacoeconomics 2006;24(7):673-682. [Medline: 16802843]

21. Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. Qual Life Res 1999;8(1-2):79-91. [Medline: 10457741]
22. Carey MP, Jorgensen RS, Weinstock RS, Sprafkin RP, Lantinga LJ, Carnrike CL, et al. Reliability and validity of the appraisal of diabetes scale. J Behav Med 1991 Feb;14(1):43-51. [Medline: 2038044]
23. Lee E, Lee YW, Lee K, Nam M, Kim YS, Han SJ. A Korean version of the Appraisal of Diabetes Scale (ADS-K): psychometric evaluation with a population of Koreans with type 2 diabetes. J Transcult Nurs 2015 May;26(3):270-278. [doi: 10.1177/1043659614524793] [Medline: 24782147]
24. Welch G, Weinger K, Anderson B, Polonsky WH. Responsiveness of the Problem Areas In Diabetes (PAID) questionnaire. Diabet Med 2003 Jan;20(1):69-72. [Medline: 12519323]
25. Bradley C. The diabetes treatment satisfaction questionnaire: DTSQ. In: Bradley C. editor. Handbook of psychology and diabetes. New York, NY: Psychology Press; 1994:111-132.
26. Bradley C. Diabetes treatment satisfaction questionnaire. Change version for use alongside status version provides appropriate solution where ceiling effects occur. Diabetes Care 1999 Mar;22(3):530-532 [FREE Full text] [Medline: 10097946]
27. Istepanian R. editor. Mobile applications (Apps) for diabetes management - Challenges and efficacy issues. 75th Annual Scientific Sessions of the American Diabetes Association 2015;Boston, MA:A.
28. Graetz I, Huang J, Muelly ER, Fireman B, Hsu J, Reed ME. Association of Mobile Patient Portal Access With Diabetes Medication Adherence and Glycemic Levels Among Adults With Diabetes. JAMA Netw Open 2020 Feb 05;3(2):e1921429 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21429] [Medline: 32074289]
29. Wu X, Guo X, Zhang Z. The Efficacy of Mobile Phone Apps for Lifestyle Modification in Diabetes: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
30. Berndt R, Takenga C, Preik P, Kuehn S, Berndt L, Mayer H, et al. Impact of information technology on the therapy of type-1 diabetes: a case study of children and adolescents in Germany. J Pers Med 2014;4(2):200-217. [doi: 10.3390/jpm4020200] [Medline: 25563223]
31. Sharma A, Mentz RJ, Granger BB, Heitner JF, Cooper LB, Banerjee D, et al. Utilizing mobile technologies to improve physical activity and medication adherence in patients with heart failure and diabetes mellitus: Rationale and design of the TARGET-HF-DM Trial. Am Heart J 2019 May;211:22-33. [doi: 10.1016/j.ahj.2019.01.007] [Medline: 30831331]
32. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of Lifestyle-Focused Text Messaging on Risk Factor Modification in Patients With Coronary Heart Disease: A Randomized Clinical Trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
33. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile Telephone Text Messaging for Medication Adherence in Chronic Disease: A Meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
34. Marcolino MS, Maia JX, Alkmim MBM, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]
35. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]
36. Lee MJ, Cha J, Choi HA, Woo S, Kim S, Wang S, et al. Blood-brain barrier breakdown in reversible cerebral vasoconstriction syndrome: Implications for pathophysiology and diagnosis. Ann Neurol 2017 Mar;81(3):454-466. [doi: 10.1002/ana.24891] [Medline: 28195428]
37. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
38. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
39. Ishikawa M, Takai M, Maeda H, Kanamori A, Kubota A, Amemiya H, et al. Factors Predicting Therapeutic Efficacy of Combination Treatment With Sitagliptin and Insulin in Type 2 Diabetic Patients: The ASSIST-K Study. J Clin Med Res 2015 Aug;7(8):607-612 [FREE Full text] [doi: 10.14740/jocmr2149w] [Medline: 26124906]
40. Knapp S, Manroa P, Doshi K. Self-monitoring of blood glucose: Advice for providers and patients. Cleve Clin J Med 2016 May;83(5):355-360 [FREE Full text] [doi: 10.3949/ccjm.83a.14147] [Medline: 27168511]

Abbreviations

ADDQOL: Audit of Diabetes Dependent Quality of Life
ADS: Appraisal of Diabetes Scale

DTSQ: Diabetes Treatment Satisfaction Questionnaire
HbA_{1c}: glycated hemoglobin
mHealth: mobile health care
PAID: Problem Areas in Diabetes
SDSCA: Summary of Diabetes Self-Care Activities

Edited by G Eysenbach; submitted 06.04.20; peer-reviewed by M Peeples, C Reis, C Liao, E Franco; comments to author 28.04.20; revised version received 07.06.20; accepted 11.08.20; published 18.09.20.

Please cite as:
Lee DY, Yoo SH, Min KP, Park CY
Effect of Voluntary Participation on Mobile Health Care in Diabetes Management: Randomized Controlled Open-Label Trial
JMIR Mhealth Uhealth 2020;8(9):e19153
URL: http://mhealth.jmir.org/2020/9/e19153/
doi:10.2196/19153
PMID:32945775

Original Paper

Checklists for Complications During Systemic Cancer Treatment Shared by Patients, Friends, and Health Care Professionals: Prospective Interventional Cohort Study

Helen V Jones[1*], MD; Harry Smith[2], MD; Tim Cooksley[3], MD; Philippa Jones[4*]; Toby Woolley[1], MD; Derick Gwyn Murdoch[5*]; Dafydd Thomas[6]; Betty Foster[6]; Valerie Wakefield[6]; Pasquale Innominato[1,7,8*], MD, PhD; Anna Mullard[1*], MD, MBBCh, MRCP; Niladri Ghosal[9], MD; Christian Subbe[10], MD, FRCP

[1]Ysbyty Gwynedd, Penrhosgarnedd, Bangor, United Kingdom
[2]School of Medicine, Cardiff Univeristy, Cardiff, United Kingdom
[3]The Christie, Manchester, United Kingdom
[4]UK Oncology Nursing Society, Marlow, United Kingdom
[5]Galactig, Lôn Cae Ffynnon Cibyn, Caernarfon, United Kingdom
[6]North Wales Cancer Forum, Bangor, United Kingdom
[7]Cancer Chronotherapy Team, Warwick Medical School, Coventry, United Kingdom
[8]European Laboratory U935, Institut National de la Santé et de la Recherche Médicale (INSERM), Paris-Saclay University, Villejuif, France
[9]North Wales Cancer Centre, Rhyl, United Kingdom
[10]School of Medical Sciences, Bangor University, Bangor, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Christian Subbe, MD, FRCP
School of Medical Sciences
Bangor University
Brigantia Building
Bangor, LL57 2AS
United Kingdom
Phone: 44 +447771922890
Email: csubbe@hotmail.com

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2020/10/e24816/

Abstract

Background: Advances in cancer management have been associated with an increased incidence of emergency presentations with disease- or treatment-related complications.

Objective: This study aimed to measure the ability of patients and members of their social network to complete checklists for complications of systemic treatment for cancer and examine the impact on patient-centered and health-economic outcomes.

Methods: A prospective interventional cohort study was performed to assess the impact of a smartphone app used by patients undergoing systemic cancer therapy and members of their network to monitor for common complications. The app was used by patients, a nominated "safety buddy," and acute oncology services. The control group was made up of patients from the same institution. Measures were based on process (completion of checklists over 60 days), patient experience outcomes (Hospital Anxiety and Depression Scale and the General version of the Functional Assessment of Cancer Therapy at baseline, 1 month, and 2 months) and health-economic outcomes (usage of appointments in primary care and elective and unscheduled hospital admissions).

Results: At the conclusion of the study, 50 patients had completed 2882 checklists, and their 50 "safety buddies" had completed 318 checklists. Near daily usage was maintained over the 60-day study period. When compared to a cohort of 50 patients with matching disease profiles from the same institution, patients in the intervention group had comparable changes in Hospital Anxiety

and Depression Scale and General version of the Functional Assessment of Cancer Therapy. Patients in the Intervention Group required a third (32 vs 97 nights) of the hospital days with overnight stay compared to patients in the Control Group, though the difference was not significant. The question, "I feel safer with the checklist," received a mean score of 4.27 (SD 0.87) on a Likert scale (1-5) for patients and 4.55 (SD 0.65) for family and friends.

Conclusions: Patients undergoing treatment for cancer and their close contacts can complete checklists for common complications of systemic treatments and take an active role in systems supporting their own safety. A larger sample size will be needed to assess the impact on clinical outcomes and health economics.

(JMIR Mhealth Uhealth 2020;8(9):e19225) doi:10.2196/19225

KEYWORDS

cancer; patient safety; checklist; quality of life; anxiety; depression; health economics; mHealth; smartphone; redundancy

Introduction

Advances in cancer management continue to improve patient outcomes but are also associated with an increase in emergency presentations with disease- or treatment-related complications [1]. The challenges of acute oncology presentations have led to an interest in developing optimal care models and support systems for meeting patients' needs [2]. Cancer patients seeking emergency care generally have longer lengths of stay, higher admission rates, and higher mortality than non-cancer patients [3].

Individualized management of acute cancer presentations is important to ensure services can mirror routine cancer care [4]. There is an increasing number of acute cancer presentations that can be risk-assessed for care in an outpatient ambulatory setting utilizing technology to support clinicians and patients. Complications of cancer and its treatments are predictable (fever, diarrhea, skin reactions, and drug-specific effects) and, in part, preventable.

Patients, friends, family, and other carers are often able to identify deviations from a patient's normal status as a first step to facilitate calls for help. Peer support has been used in other settings to improve clinical care and safety, allowing families and friends to look after vulnerable patients, including those discharged after a stroke [5]. Mobile health apps for patients with cancer have the potential to track deterioration [6], support education, and recovery [7-9].

Modular redundancy is the duplication of critical components of a system to increase the reliability of performance in the design of technology [10] or clinical services [11]. Checklists allow redundancy by allowing multiple users to verify safety and are widely used in health care [12-15]. The United Kingdom Oncology Nursing Society (UKONS) has developed checklists for symptom-driven telephone triage [16].

Patients are competent to carry out surveillance and management of chronic conditions, as demonstrated by people with diabetes checking their blood sugars, people with asthma monitoring their peak flows, and people with heart failure recording their weight. Patients admitted to the hospital as medical emergencies can assist in the recording of key safety-critical information [17]. Information about cancer improves compliance, especially if it is tailored to individual needs and context-specific [4].

The study aimed to test the feasibility of a smartphone-based checklist that allows redundant access to safety-critical processes for patients, members of their immediate social network, and health care professionals to stimulate patients and carers to seek medical assistance when necessary while providing reassurance when appropriate.

Methods

Study Design

The trial was designed as a prospective interventional cohort study.

Participants

Oncology patients attending outpatient clinics at the Ysbyty Gwynedd, North Wales, were invited to take part in the study. Patients were eligible if they had a known malignancy and were receiving treatment for cancer, including chemotherapy, radiotherapy, immunotherapy, or best supportive care. Patients were eligible for enrollment during the entirety of their treatment course.

Patients were excluded from the trial if they were receiving end of life care or lacked a smartphone to access the app. There were 50 app licenses available for the trial, and 100 patients were recruited. Patients who did not want to use the smartphone app and patients recruited after all the licenses had been distributed were recruited into a control group to provide indicative data on service usage in patients not using the app.

All participants, including controls, gave written informed consent.

Smartphone App

The content of the app was coproduced in four focus group events. Focus groups consisted of 15 patient representatives, clinicians, and health-service researchers. Checklists were based on the UKONS 24-hour triage tool [16], the UKONS Oncology/Hematology risk assessment tool for Primary Healthcare Professionals [18], and a symptom assessment tool included in the Cancer Research UK Patient treatment record adapted from the UKONS 24-Hour Triage Tool [19]. UKONS tools classify symptoms and signs according to risk and urgency into green, amber, and red with linked actions for escalation from generic advice (green) to encouragement to seek a routine appointment or an urgent assessment (amber or red) (Figure 1).

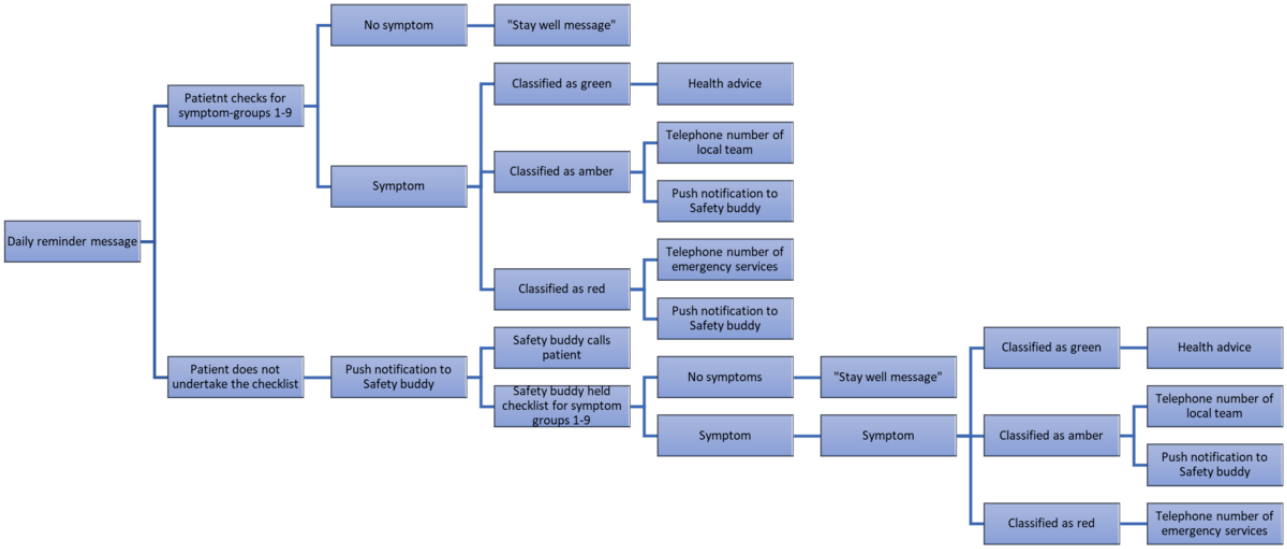
Figure 1. Sample screenshots of checklist items. Item 2 (breathlessness or chest pain) is linked to a red escalation, item 7 (urine problems) is linked to amber escalation.

Clinicians, patients, and researchers devised a hierarchy of safety-relevant symptoms in two iterations. Items were summarized in nine screens, and item rankings were decided by consensus. Checks were presented in order of priority, starting with the most urgent and life-threatening symptoms.

The system allows the addition of disease- or treatment-specific checks in the content management system. For this study, only a generic checklist was activated. Prototypes were tested against typical case studies. Symptoms that were "red flags" generated a recommendation to the patient to seek medical care. The app sent text reminders to patients once a day to complete the checklist.

Each patient was asked to invite one family member or friend to be their "Safety buddy." Safety buddies also downloaded the checklist app to their smartphone. Safety buddies received push notifications if the patient did not complete the checklist within an agreed timeframe or if patients reported potentially serious symptoms (equivalent to red fields in the UKONS checklist): "You might want to call your friend/family member." Safety buddies were then asked to complete the checklist on their phone with the patient (Figure 2).

Figure 2. Workflow of the application.



A dashboard for the acute oncology team showed notifications and alerts that could be annotated by clinicians. Nurse specialists reviewed the reported symptoms once daily via an online dashboard and followed up with patients if the symptoms required further attention.

Use of the App

Patients were enrolled for 60 days. Patients and the friend or family member were encouraged to access the app to record symptoms at least once daily. It was emphasized to patients that nursing staff would not be monitoring the app constantly and, therefore, the onus was on them to seek medical care if urged to do so by the app. App users received a call after a week to check for technical difficulties in app usage.

Outcome Measures

The Hospital Anxiety and Depression Scale (HADS) [20] and the General version of the Functional Assessment of Cancer Therapy (FACT-G) [21] were completed at baseline, one month, and two months.

Health-economic outcomes consisted of the usage of appointments in primary care and elective and unscheduled hospital admissions.

Patient Feedback

Patients and carers were able to provide feedback within the application. They were asked to use a Likert scale with gradings from 1 (strongly disagree) to 5 (strongly agree).

Project Governance

The study was conducted according to the principles of the World Medical Association's Declaration of Helsinki 2013 [22]. A study board supervised the development, testing, and evaluation. The group met every three months to issue interim reports and to review risk logs and possible adverse events. Ethics approval was granted (REC reference: 18/WA/0213).

Results

Recruitment

Patients were recruited from January 24, 2019, to September 17, 2019. Of the 197 patients approached, 100 agreed to participate—50 in the control group and 50 in the intervention group. Of the 100 participants, 56 were female. Groups were matched for gender, type of cancer, and performance status, but patients in the control group were older (mean 59, SD 13 years vs mean 68 SD 13 years; $P<.001$) (Table 1).

Table 1. Participants, comorbidities, cancer type, performance status, and treatment.

Item	Intervention	Control	*P* value[a]
Age (years), mean (SD)	59 (13)	68 (13)	<.001
Gender female, n (%)	27 (54)	29 (58)	.69
Comorbidities, n			__[b]
Diabetes	2	8	
Chronic obstructive pulmonary disease	4	5	
Ischemic heart disease	1	5	
Cancer type, n (%)			.36
Breast	12 (24)	11 (22)	
Bowel	13 (26)	9 (18)	
Lung	6 (12)	15 (30)	
Kidney	3 (6)	2 (4)	
Rectal	3 (6)	2 (3)	
Pancreatic	1 (2)	3 (6)	
Prostate	2 (4)	2 (4)	
Esophagus	2 (4)	2 (4)	
Testicular	3 (6)	0 (0)	
Ovarian	1 (2)	2 (4)	
Rectal	3 (6)	0 (0)	
Endometrial	2 (4)	0 (0)	
Gastric	1 (2)	0 (0)	
Leiomyosarcoma	0 (0)	1 (2)	
Mesothelioma	0 (0)	1 (2)	
Performance status, n			.31
0	20	16	
1	22	21	
2	5	9	
3	0	2	
Treatment, n			.67
Chemotherapy	45	45	
Radio- and chemotherapy	3	2	
Surgery and chemotherapy	1	1	
Surgery	0	1	
Best supportive care	1	1	

[a]Chi-square test.

[b]Not applicable.

Checklist Utilization

Checklists were used 2882 times by the 50 patients in the intervention group, a median of 62 times per patient with the number of uses ranging from 13 to 102 times over the study period. App use resulted in no alert being generated on 2715 occasions, indicating no or no significant symptoms. On 167 (5.8%) occasions, actions were advised. There were 130 green alerts, 28 amber alerts, and 9 red alerts.

Usage by patients was 284 times in the first week, 347 times in the second week, and 228 times in the ninth week of participation.

Of the 50 nominated friends and family members, 31 used the checklists in the app to support their patient partner for a total of 318 times. Usage generated no alert on 267 occasions; in 28 instances, contact of a health care professional was advised. There were 18 amber alerts, 9 red, and 1 major alert.

Friends and family members used the app 77 times in the first week, 67 times in the second week, and 16 times in the ninth week.

Symptoms flagged by the checklists were, in order of frequency, exhaustion (102), nausea (23), fever (14), chest pain (13), sore mouth (13), diarrhea (11), pain (8), constipation (5), skin and eye complaints (4), pins and needles (3), mental health issues (2), visual disturbances (1), and urinary symptoms (1).

Logs completed by the acute oncology team indicated 23 patient calls in response to checklist items. Calls covered a broad range of topics, including technical advice (2 calls), reassurance (8 calls), and advice to admit (4 calls).

Clinical Outcomes

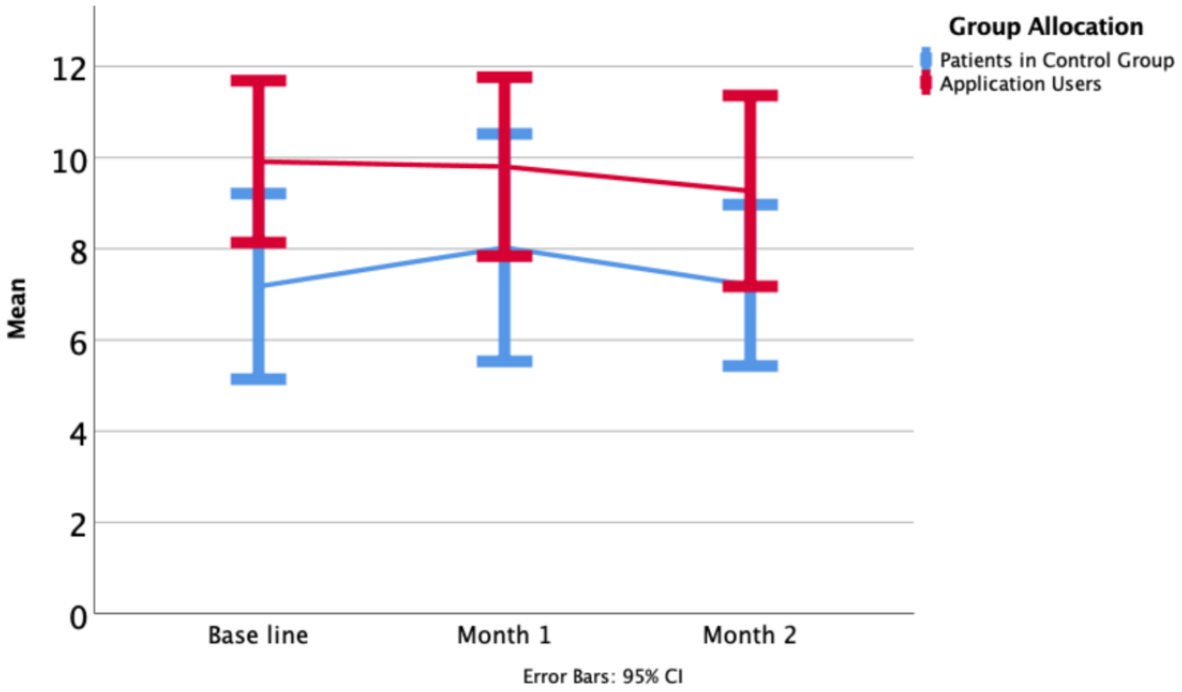
Patients in the intervention group had 19 scheduled inpatient days, 40 unscheduled days in the hospital, and 32 unscheduled days in the hospital involving overnight stays. Patients in the control group had 2 scheduled inpatient days, 108 unscheduled days in the hospital, and 97 unscheduled days in the hospital involving overnight stays. There were 40 patients in the intervention group and 38 patients in the control group who spent no unscheduled time in the hospital. Patients in the intervention group required a third as many hospital days with overnight stay in comparison to the control group.

Patients and primary care practices requested information about appointments in primary care. In the first week after enrollment, patients in the intervention group saw their general practitioner 10 times, and patients from the control group 3 times. In the subsequent 3 weeks, patients from the intervention group saw a general practitioner 20 times, and patients in the control group saw a general practitioner 14 times. In the second month, patients from the intervention group saw a general practitioner 30 times, and patients from the control group had 15 visits.

Anxiety and Depression

Patient experience was captured by standardized questionnaires and informal feedback from inside the application. A HADS score of 11 or more indicates clinically significant anxiety or depression. At baseline, the average HADS score was 7.9 (SD 7.2) in the control group and 10.2 (SD 6.1) in the intervention group (Figure 3). HADS scores of 11 or greater were observed in 14 patients in the control group and 26 patients in the intervention group. After one month, participants with a HADS score >11 had declined to 13 in the control group and 20 in the intervention group. At two months, 11 control patients and 18 intervention patients fulfilled the same criteria.

Figure 3. Hospital Anxiety & Depression Scale mean at baseline, one month, and at the end of the study period.



Values for the FACT-G were not significantly different at baseline, one month, or two months. Over the full duration of the study, 6 patients in the control group and 10 patients in the intervention group improved by more than 10% over their baseline (Figures 4-6).

Figure 4. General version of the Functional Assessment of Cancer Therapy (FACT-G) overall means at baseline, one month, and at the end of the study period.

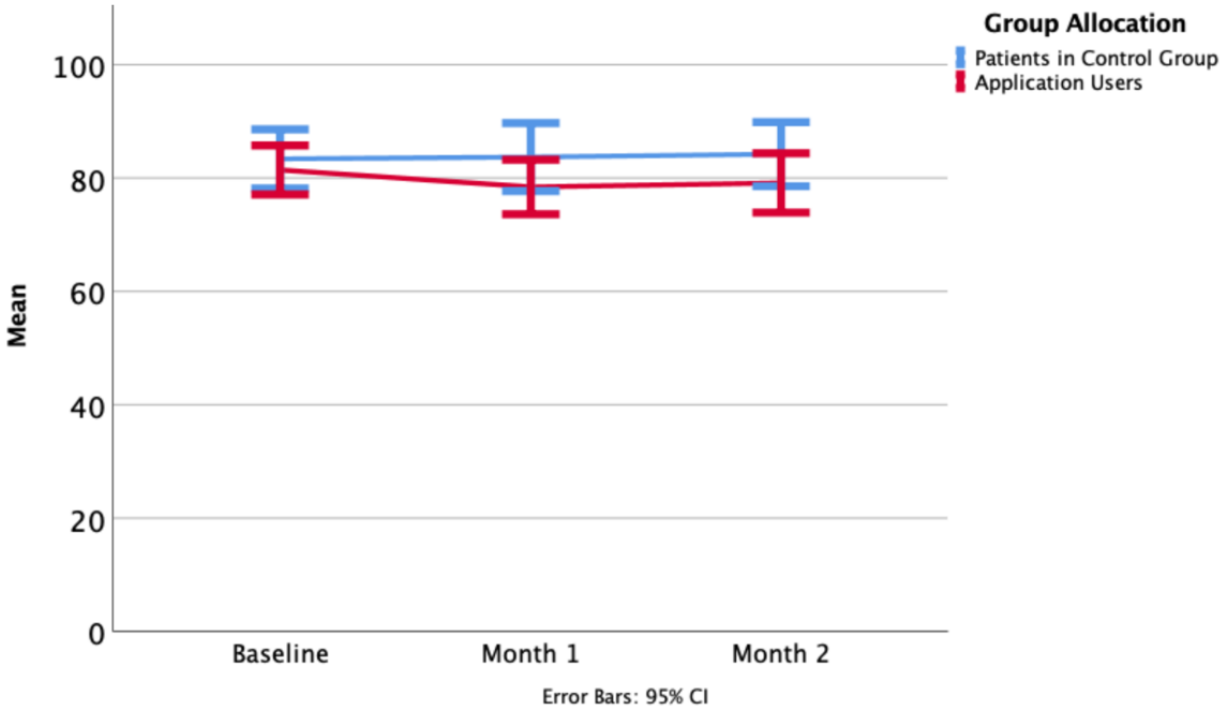


Figure 5. General version of the Functional Assessment of Cancer Therapy (FACT-G) Social and Family wellbeing subscore means at baseline, one month, and at the end of the study period.

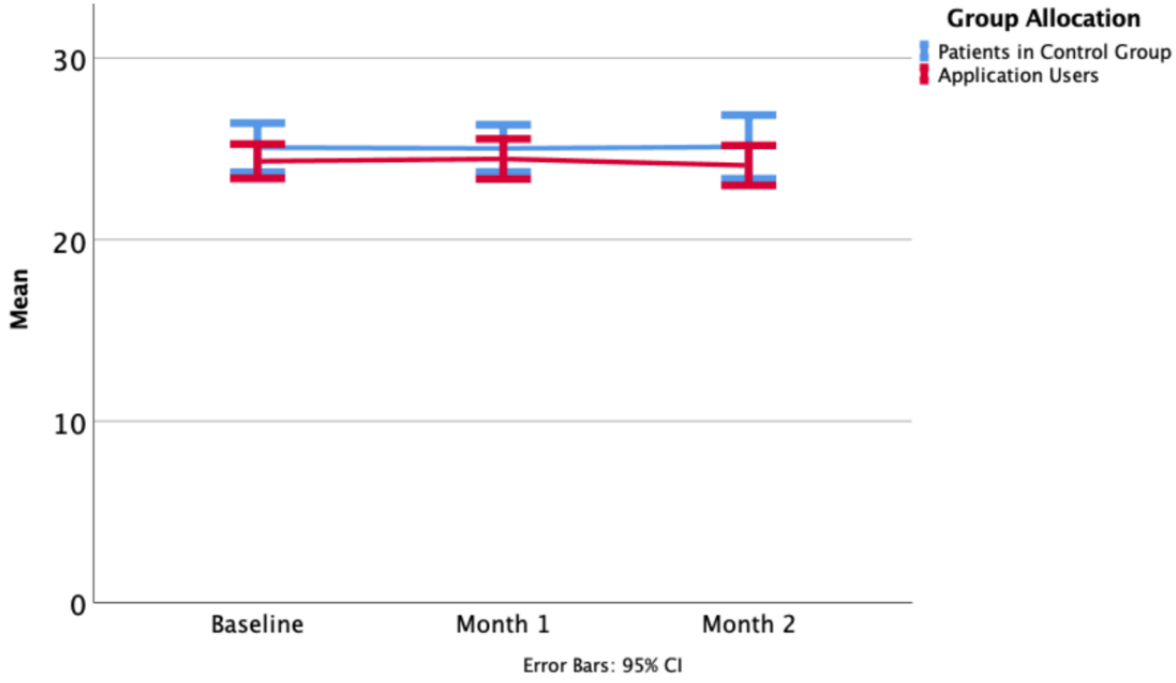
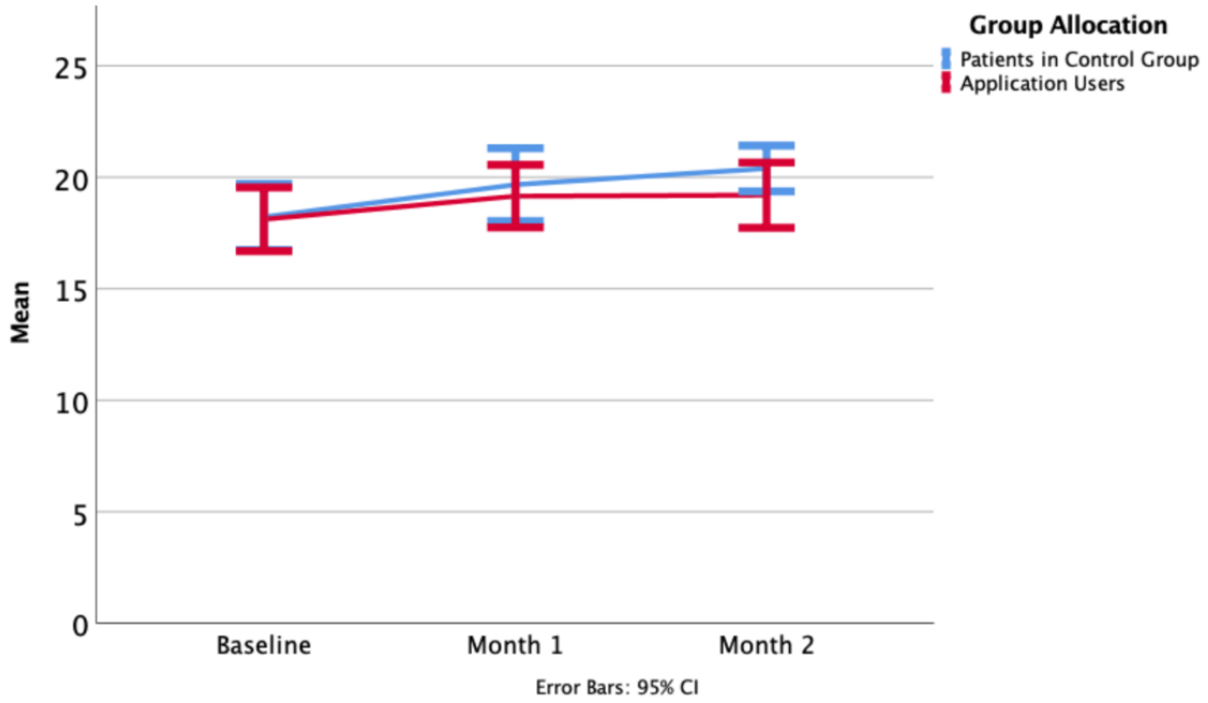


Figure 6. General version of the Functional Assessment of Cancer Therapy (FACT-G) Emotional wellbeing subscore means at baseline, one month, and at the end of the study period.



Feedback From Patients, Friends, and Families

Structured feedback was submitted from within the application by 48 patients and 25 family members and friends (Table 2). Mean ratings on a Likert scale with values from 1 (strongly disagree) to 5 (strongly agree) to the question "I feel safer with the checklist" was 4.27 (SD 0.87) for patients and 4.55 (SD 0.65) for family members and friends. The question "The link to a health care professional is helpful" yielded mean ratings of 4.61 (SD 0.74) for patients and 4.75 (SD 1.35) for family members and friends.

Table 2. Structured feedback using Likert scales with values from 1 (strongly disagree) to 5 (strongly agree).

Feedback question	Patients (all)	Family and friends (all)	Patients		
			First assessment	Last assessment	Delta
The information in the app is helpful, mean	4.32	4.2	4.07	4.45	0.38
The link to a friend or relative is helpful, mean	4.48	4.41	4.42	4.27	–0.15
The link to a health care professional is helpful, mean	4.61	4.75	4.55	4.66	0.11
I feel safer with the checklist, mean	4.27	4.55	4.17	4.41	0.24

Discussion

Principal Findings

With the KeepMeSafe application, patients and their families and friends were able to use a smartphone app to work through a list of common complications of cancer and systemic therapies. We demonstrated the feasibility of assistance by members of the patients' social network at times when patients felt unable to complete the checklist themselves. To our knowledge, this is the first time that patients and members of their social network have been deployed as redundant parts of a safety system.

Patients were only able to participate if they owned a smartphone. This limitation might exclude some patients, but the percentage of people actively using smartphones in the UK in 2019 was 82.9%, the highest in the world [23]. The limited size of our study and the fact that the intervention and control groups were not randomized or matched means that the study does not allow conclusions about clinical outcomes, effectiveness, or efficiency.

Patients experience higher levels of anxiety and depression than the general population [22] though consistent with other contemporaneous cohorts of people who have cancer in the UK [24].

In a review of the literature of clinical trials involving mobile health apps, we found 17 studies of between 12 and 2352 patients [25]. Smartphone apps or internet portals primarily collected data on clinical symptoms or activity data with some improvement in patient-reported outcome measures. The authors

found limited evidence for effects on mortality or cancer-related morbidity, including complications and health-economic outcomes. Many studies did not report on app usage. Only a few studies have reported improvements in quality of life [26,27]. App for monitoring pain and linked to the ability to escalate to a clinician might lead to improved symptom control [28]. Recruitment rates of 50% in our study are comparable to other trials in this field [29].

We collected data of indicative health-service usage by reporting days spent in the hospital and appointments with primary care physicians. We observed trends towards increased usage of primary care appointments and decreased usage of hospital days for unscheduled admissions in the younger intervention group.

App usage was high and comparable with other high-quality applications [30]. Patients and their buddies reported satisfaction with the information in the app and its links to health care staff and reported feeling safer with the application. It is difficult to say that patients felt more empowered to reach out to health care staff (or that the app encouraged them to do so), given that the majority of contacts were initiated by nursing staff. At least eight of these contacts received telephone advice, however, and it might be inferred that having easy access to health care staff in this way reduced the burden on primary care services.

Limitations identified in the literature review were addressed by measuring app usage and validated clinical outcome measures and surrogates for health economic measures, albeit in a non-randomized single-center study. Many mobile health apps are designed for single diseases [31-33] or use generic metrics such as physical activity [34], with only a few applications reporting patient outcomes [35]. By using a content management system as the underlying architecture, we enabled agile, modular development for future expansion to rarer complications, tailoring to different cancers, individualized treatment regimes, and patient preferences. While this study was limited in scope to proof-of-concept, it has generated the methodology for larger trials powered to demonstrate improvements in patient-centric outcomes.

The study demonstrates that patients and those close to them can take an active part in a redundant safety system. Technology can facilitate laypersons to undertake some of the safety-critical screening functions that are normally undertaken by nurses based on the UKONS clinical checklists.

Real-time response to alerts would require 24/7 cover of staff who are familiar with diseases and treatment modalities. Scale-up of usage, including utilization for follow-up of patients with cancer, would require limited investment into the soft-ware platform but reliable investment into the teams that support cancer services locally and nationally.

This application may be a useful tool in aiding patients to access early and appropriate acute cancer care. It may also have a role in supporting ambulatory outpatient management of presentations suitable for this model of care.

Future research will have to tease out the effect size in multiple settings. The number of friends and family members forming a safety network for patients may be relevant for the effect-size; several safety partners might support patients better than just a single partner.

Ways to strengthen ownership and activation of patients in future versions of the application might include incentives for usage or link to continuous monitoring with wearable sensors to supplement patient-reported symptoms with quantitative measures of risk [36].

The hypothesis of the checklist application that remains to be tested in larger trials is that usage of electronic checklists tailored to the needs of patients with cancer will improve reliability and timeliness of engagement with their multi-disciplinary team.

We hope to affect patients with cancer positively by first facilitating safer care: complications are, in large part, predictable. Checklists allow patients to be actively involved in the prevention of adverse events. Modular redundancy of safety-critical processes is a key mechanism to provide safe and stable systems in other industries [10]. The usage of checklists by multiple partners should ultimately lead to a testable reduction in preventable adverse outcomes. Lastly, we believe in the value of greater autonomy of patients through participation. Access to safety-critical information in a personalized and context-specific way is key for patient activation [37]. We fully expect that this will also improve resilience to acute complications [38].

Conclusions

We coproduced a checklist application for smartphones with cancer patients, their friends, and families and demonstrated proof of concept as a networked and scalable safety intervention.

It is feasible to enable patients undergoing treatment for cancer to contribute to their own safety in recognizing complications of cancer and their therapy. To assess the impact on clinical outcomes requires larger randomized trials but utilizing such applications may form a key aspect of future acute cancer care.

Acknowledgments

Funding for the study was provided by Tenovus, and the application was developed in collaboration with Galactig (part of Rondo Media Group). The authors are most grateful for the generous support of all involved in funding, setting up, and undertaking the study. We are specifically indebted to the patient representatives, the acute oncology nurses Dawn Griffiths and Louise Preston-Jones and the research team at the Ysbyty Gwynedd, especially Laura Longshaw, Gareth Jones, Dianne Thomas, Alice Thomas, and Wendy Scrasse.

We were privileged to be guided by our knowledgeable and dedicated patient representatives Val Wakefield, Betty Foster, and Dafydd Thomas from the North Wales Cancer Forum.

Conflicts of Interest

The study was funded by an iGrant from TENOVUS UK. The corresponding author (CPS) was funded in part by an Improvement Science Fellowship from The Health Foundation.

References

1. Navani V. How has acute oncology improved care for patients? Curr Oncol 2014 Jun;21(3):147-149 [FREE Full text] [doi: 10.3747/co.21.1904] [Medline: 24940096]
2. Cooksley T, Rice T. Emergency oncology: development, current position and future direction in the USA and UK. Support Care Cancer 2017 Jan;25(1):3-7. [doi: 10.1007/s00520-016-3470-1] [Medline: 27815712]
3. Cooksley T, Kitlowski E, Haji-Michael P. Effectiveness of Modified Early Warning Score in predicting outcomes in oncology patients. QJM 2012 Nov 01;105(11):1083-1088 [FREE Full text] [doi: 10.1093/qjmed/hcs138] [Medline: 22855285]
4. Hibbard J, Helen G. Supporting people to manage their health: an introduction to patient activation. In: The King's Fund. London: The King's Fund; 2014.
5. Rogerson L, Burr J, Tyson S. The feasibility and acceptability of smart home technology using the Howz system for people with stroke. Disabil Rehabil Assist Technol 2020 Feb;15(2):148-152. [doi: 10.1080/17483107.2018.1541103] [Medline: 30663921]
6. Theile G, Klaas V, Tröster G, Guckenberger M. mHealth Technologies for Palliative Care Patients at the Interface of In-Patient to Outpatient Care: Protocol of Feasibility Study Aiming to Early Predict Deterioration of Patient's Health Status. JMIR Res Protoc 2017 Aug 16;6(8):e142 [FREE Full text] [doi: 10.2196/resprot.7676] [Medline: 28814378]
7. Geng Y, Myneni S. Patient Engagement in Cancer Survivorship Care through mHealth: A Consumer-centered Review of Existing Mobile Applications. In: AMIA Annu Symp Proc. Patient Engagement in Cancer Survivorship Care through mHealth: A Consumer-centered Review of Existing Mobile Applications. AMIA . Annu Symp proceedings AMIA Symp. 2015; 2015 Presented at: AMIA 2015 Annual Symposium; Nov. 14-18; San Francisco p. 580-588.
8. Davis SW, Oakley-Girvan I. mHealth Education Applications Along the Cancer Continuum. J Cancer Educ 2015 Jun;30(2):388-394 [FREE Full text] [doi: 10.1007/s13187-014-0761-4] [Medline: 25482319]
9. Nasi G, Cucciniello M, Guerrazzi C. The role of mobile technologies in health care processes: the case of cancer supportive care. J Med Internet Res 2015 Feb 12;17(2):e26 [FREE Full text] [doi: 10.2196/jmir.3757] [Medline: 25679446]
10. Wakerly J. Microcomputer reliability improvement using triple-modular redundancy. New York; 1976 Jun Presented at: Proceedings of the IEEE; 1976; New York p. 889-895 URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a042246.pdf [doi: 10.1109/proc.1976.10239]
11. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
12. Daniels R, Nutbeam T, McNamara G, Galvin C. The sepsis six and the severe sepsis resuscitation bundle: a prospective observational cohort study. Emerg Med J 2011 Jun;28(6):507-512. [doi: 10.1136/emj.2010.095067] [Medline: 21036796]
13. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AHS, Dellinger EP, Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med 2009 Jan 29;360(5):491-499. [doi: 10.1056/NEJMsa0810119] [Medline: 19144931]
14. Arriaga AF, Bader AM, Wong JM, Lipsitz SR, Berry WR, Ziewacz JE, et al. Simulation-based trial of surgical-crisis checklists. N Engl J Med 2013 Jan 17;368(3):246-253. [doi: 10.1056/NEJMsa1204720] [Medline: 23323901]
15. Subbe CP, Kellett J, Barach P, Chaloner C, Cleaver H, Cooksley T, Crisis Checklist Collaborative. Crisis checklists for in-hospital emergencies: expert consensus, simulation testing and recommendations for a template determined by a multi-institutional and multi-disciplinary learning collaborative. BMC Health Serv Res 2017 May 08;17(1):334 [FREE Full text] [doi: 10.1186/s12913-017-2288-y] [Medline: 28482890]
16. Oncology/Haematology 24 Hour Triage Rapid Assessment and Access Toolkit. https://www.ukons.org/site/assets/files/1134/oncology_haematology_24_hour_triage.pdf. London: UKONS; 2016. URL: https://www.ukons.org/site/assets/files/1134/oncology_haematology_24_hour_triage.pdf [accessed 2020-09-05]
17. Subbe C. Driving safety and signposting risk - Report to the SHINE programme of the Health Foundation. Bangor. The Health Foundation. London: The Health Foundation; 2015. URL: https://www.health.org.uk/sites/default/files/Betsi%20Cadwaladr%20final%20report_website%20version.pdf [accessed 2020-09-05]
18. United Kingdom Oncology Nursing Society. Oncology/Haematology Treatment Toxicity Risk Assessment Tool For Primary Healthcare Professionals. London. London: UKONS; 2017. URL: https://www.macmillan.org.uk/_images/oncology-treatment-toxicity-risk-assesment-tool_tcm9-317392.pdf [accessed 2020-09-05]
19. Your Cancer Treatment Record. London: Cancer Research UK; 2018. URL: https://publications.cancerresearchuk.org/publication/your-cancer-treatment-record [accessed 2020-09-05]
20. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]

21. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993 Mar;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
22. Breeman S, Cotton S, Fielding S, Jones GT. Normative data for the Hospital Anxiety and Depression Scale. Qual Life Res 2015 Feb;24(2):391-398. [doi: 10.1007/s11136-014-0763-z] [Medline: 25064207]
23. Global Mobile Market Report 2019 - Light Version. Newzoo.com.: newzoo; 2019. URL: https://resources.newzoo.com/hubfs/Reports/2019_Free_Global_Mobile_Market_Report.pdf [accessed 2020-09-05]
24. Lindner OC, McCabe MG, Boele F, Mayes A, Talmi D, Radford J, et al. Discussing factors associated with quality of life in cancer follow-up appointments: a preliminary test of a pragmatic model for clinical practice. Clin Rehabil 2019 Apr;33(4):762-772. [doi: 10.1177/0269215518820093] [Medline: 30582361]
25. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? A critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar;28(3):1469-1479 [FREE Full text] [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]
26. Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile Phone Apps for Quality of Life and Well-Being Assessment in Breast and Prostate Cancer Patients: Systematic Review. JMIR Mhealth Uhealth 2017 Dec 04;5(12):e187 [FREE Full text] [doi: 10.2196/mhealth.8741] [Medline: 29203459]
27. Fjell M, Langius-Eklöf A, Nilsson M, Wengström Y, Sundberg K. Reduced symptom burden with the support of an interactive app during neoadjuvant chemotherapy for breast cancer - A randomized controlled trial. Breast 2020 Jun;51:85-93 [FREE Full text] [doi: 10.1016/j.breast.2020.03.004] [Medline: 32247251]
28. Zheng C, Chen X, Weng L, Guo L, Xu H, Lin M, et al. Benefits of Mobile Apps for Cancer Pain Management: Systematic Review. JMIR Mhealth Uhealth 2020 Jan 23;8(1):e17055 [FREE Full text] [doi: 10.2196/17055] [Medline: 32012088]
29. Gustavell T, Sundberg K, Langius-Eklöf A. Using an Interactive App for Symptom Reporting and Management Following Pancreatic Cancer Surgery to Facilitate Person-Centered Care: Descriptive Study. JMIR Mhealth Uhealth 2020 Jun 17;8(6):e17855 [FREE Full text] [doi: 10.2196/17855] [Medline: 32554375]
30. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Hälleberg Nyman M, Wengström Y, et al. Adherence to Report and Patient Perception of an Interactive App for Managing Symptoms During Radiotherapy for Prostate Cancer: Descriptive Study of Logged and Interview Data. JMIR Cancer 2017 Oct 31;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
31. Denis F, Viger L, Charron A, Voog E, Dupuis O, Pointreau Y, et al. Detection of lung cancer relapse using self-reported symptoms transmitted via an internet web-application: pilot study of the sentinel follow-up. Support Care Cancer 2014 Jun;22(6):1467-1473. [doi: 10.1007/s00520-013-2111-1] [Medline: 24414998]
32. Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast Cancer: A randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018 Aug;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]
33. Golsteijn RHJ, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Oct 30;15(1):106 [FREE Full text] [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
34. Kanera IM, Willems RA, Bolman CAW, Mesters I, Verboon P, Lechner L. Long-term effects of a web-based cancer aftercare intervention on moderate physical activity and vegetable consumption among early cancer survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2017 Feb 10;14(1):19 [FREE Full text] [doi: 10.1186/s12966-017-0474-2] [Medline: 28187725]
35. Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, et al. Electronic Systems for Patients to Report and Manage Side Effects of Cancer Treatment: Systematic Review. J Med Internet Res 2019 Jan 24;21(1):e10875 [FREE Full text] [doi: 10.2196/10875] [Medline: 30679145]
36. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The Future of Wearable Technologies and Remote Monitoring in Health Care. Am Soc Clin Oncol Educ Book 2019 Jan;39:115-121 [FREE Full text] [doi: 10.1200/EDBK_238919] [Medline: 31099626]
37. 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. [doi: 10.1136/amiajnl-2013-002141] [Medline: 24272163]
38. Bradley-Roberts E, Subbe C. Role of Psychological Resilience on Health-Outcomes in Hospitalized Patients with Acute Illness: A Scoping Review. Acute Med 2017;16(1):10-15. [Medline: 28424799]

Abbreviations

FACT-G: General version of the Functional Assessment of Cancer Therapy
UKONS: United Kingdom Oncology Nursing Society

Edited by G Eysenbach; submitted 15.04.20; peer-reviewed by E Marshall, Z Zhang; comments to author 09.07.20; revised version received 09.08.20; accepted 18.08.20; published 25.09.20.

Please cite as:
Jones HV, Smith H, Cooksley T, Jones P, Woolley T, Gwyn Murdoch D, Thomas D, Foster B, Wakefield V, Innominato P, Mullard A, Ghosal N, Subbe C
Checklists for Complications During Systemic Cancer Treatment Shared by Patients, Friends, and Health Care Professionals: Prospective Interventional Cohort Study
JMIR Mhealth Uhealth 2020;8(9):e19225
URL: http://mhealth.jmir.org/2020/9/e19225/
doi:10.2196/19225
PMID:32975526

Original Paper

Mobile Health App for Self-Learning on HIV Prevention Knowledge and Services Among a Young Indonesian Key Population: Cohort Study

Priyanka Rani Garg[1], PhD; Leena Uppal[1], MA; Sunil Mehra[1], MD; Devika Mehra[1,2], PhD

[1]MAMTA Health Institute for Mother and Child, New Delhi, India

[2]Medeon Science Park, Malmo, Sweden

Corresponding Author:
Leena Uppal, MA
MAMTA Health Institute for Mother and Child
B-5, Greater Kailash Enclave-II
New Delhi, 110048
India
Phone: 91 11 29220210
Fax: 91 11 29220575
Email: leena@mamtahimc.org

Abstract

Background: Indonesia is the only country in the Asia Pacific region where the incidence of HIV is still on the rise, and its prevalence is extremely high among the key populations such as men who have sex with men, transgender women, and people who inject/use drugs. Mobile health (mHealth) apps provide an innovative platform for delivering tailored HIV prevention and care among these populations more efficiently than possible through the direct face-to-face approach.

Objective: The aim of this study was to assess the role of a peer-customized mobile app based on the principle of self-learning for improving HIV prevention knowledge and access to health services among men who have sex with men, transgender women (known as Waria in Indonesia), and people who use drugs in Indonesia.

Methods: A prospective intervention cohort study was conducted among the key populations in five provinces of Indonesia (Jakarata, West Java, East Java, Special Region of Yogyakarta, and Bali). The data were evaluated using a pre-post assessment survey conducted on a sample of 200 unique users, including 50 men who have sex with men and transgender women each, and 100 people who use drugs, with a follow-up response rate of 98% and 70%, respectively. An mHealth app named RUMAH SELA was developed and implemented among the key populations.

Results: From baseline to the endpoint of the study, there was a significant increase in comprehensive HIV-related knowledge from 20% (10/49) to 60% (29/49), 22% (11/49) to 57% (28/49), and 49% (34/70) to 74% (52/70) among men who have sex with men (P=.004), transgender women (P<.001), and people who use drugs (P<.001), respectively. There was also a reduction in sexual activities without condom use from 22% (11/49) to 19% (9/49), 18% (9/49) to 12% (6/49), and 21% (15/70) to 10% (7/70) among men who have sex with men (P=.45), transgender women (P=.25), and people who use drugs (P<.001), respectively. There was an uptake of HIV testing by 31% (15/49) for men who have sex with men, 49% (24/49) for transgender women, and 26% (18/70) for people who use drugs after using the app. There was a reduction in injecting drugs with a used needle in drug users from 45/70 (78%) to 15/70 (26%). Measures of self-esteem increased among men who have sex with men (mean 26.4 vs mean 27.1), transgender women (mean 26.5 vs mean 27.8; P=.02), and people who use drugs (mean 24.0 vs mean 25.0). In addition, 27% (7/24) of men who have sex with men, 25% (4/15) of transgender women, and 11% (2/18) of drug users made an appointment for an HIV test through the app. The app was quite highly accepted by the key populations as nearly a quarter felt that they became more confident in discussing issues about sexuality, more than 80% found that the app provided sufficient knowledge about HIV, and more than half of the participants found the app to be user friendly.

Conclusions: This one-of-a-kind mHealth intervention with an mHealth app as a self-learning tool is effective in increasing HIV-related knowledge and behavior, and access to services with strong acceptability by the community. There is a need to scale up such interventions for efficacy testing in a larger population to provide evidence for national-level mHealth programs addressing HIV.

(JMIR Mhealth Uhealth 2020;8(9):e17646) doi:10.2196/17646

KEYWORDS

mHealth; Indonesia; HIV; key populations

Introduction

Indonesia, situated in the southeast Asia region, is witnessing an upward trend in HIV/AIDS prevalence [1]. Indonesia is the only country in which the HIV epidemic is still on the rise among the cluster of five high-burden countries (India, Myanmar, Nepal, Thailand, and Indonesia) of the southeast Asia region [2]. The national HIV prevalence in Indonesia (among people aged 15 years and above) is 0.3% in the age group 15-49 years [3]. The HIV epidemic in Indonesia is mostly concentrated among key populations, including sex workers, men who have sex with men (MSM), and transgender women (known as *Waria* in Indonesia), although the prevalence varies across provinces [4]. The epidemic is on the rise particularly among MSM and their sexual partners. The HIV prevalence among MSM in Indonesia is 12.8% and is 7.4% for transgender women [5]. Among people who inject drugs, there has been a substantial decline in HIV prevalence from 52.4% in 2007 to 28.8% in 2015, although the absolute numbers are still quite high [6].

In 2012, the Indonesian government responded to the growing HIV epidemic with the continuum of care known as *Layanan Komprehensif Berkesinambungan*, which aimed at integrating prevention, care, and treatment for all, and directed individuals toward immediate antiretroviral therapy (ART) after receiving an HIV diagnosis; however, the coverage of prevention services among key populations remained less than 55% and only 10-20% of people living with HIV were receiving ART at the end of 2016 [5,6]. Current evidence indicates that less than half of the key populations with HIV in the Asia-Pacific region, including Indonesia, are aware of their HIV status and only around a third are referred for treatment [5]. Therefore, there is a pressing need to fill these gaps and move toward innovative and cost-effective solutions to address the health care needs for these populations.

The accessibility and potential of mobile health (mHealth) or electronic health are rapidly growing, and is considered a promising approach to reach young and key populations (including MSM, transgender women, and sex workers) for addressing health care delivery gaps in the Asian region [7-10]. Indonesia ranks among the top 10 countries in the mobile connectivity index. The 3G and 4G coverage has reached over 90% and 80%, respectively, with more than half of the country's adult population now using mobile internet [11]. Encouragingly, there is evidence that online or mobile-based interventions are effective in cisgender young people, and are emerging for sexual minority youth as well; however, there are limited evidence-based interventions using mobile-based apps designed specifically for gender and sexual minority young people. Further, gamification has been viewed as a promising approach to enhance participant interaction, increase behavior change learning opportunities, and improve intervention appeal for adolescents and youth [12].

The aim of this study was to assess the role of a peer-customized mobile app based on the principle of self-learning in increasing knowledge on HIV transmission, creating awareness about safe sex and HIV prevention, facilitating behavioral change, and fostering the uptake of HIV testing for improving access to health services among the key populations, particularly among MSM, transgender women, and people who use drugs (PWUD), in Indonesia.

Methods

Participants

Participants were recruited from five provinces of Indonesia: Jakarta, West Java, East Java, Special Region of Yogyakarta, and Bali. Respondents were recruited with the help of peer leaders, who are community leaders/outreach workers associated with the program implementation organization. Peer leaders listed the key populations from their networks and mapped areas with hotspots. To qualify for the intervention, there were certain inclusion criteria defined for enrolling the participants: (1) aged between 16 and 30 years (identified by the Indonesian government as young adults) [13]; (2) identify themselves as MSM, transgender woman, or PWUD, including the use of cocaine, methamphetamine, or ecstasy in the last 12 months or in the last 4 weeks; (3) mobile phone literate and own an Android phone. A total of 200 unique users (50 MSM, 50 transgender women, and 100 PWUD) were identified to be eligible for the study and provided written informed consent (assent for those younger than 18 years). Consent was obtained from the guardians/caretaker/outreach worker for providing services to the minor participants. For each province, six peer leaders were identified who were responsible for mapping the app users, conducting the study, and monitoring the use of the app during the study period.

Pre-Post Assessment

A baseline survey was carried out to assess the knowledge about HIV and sexual behavior of the participants in August 2017, followed by an endpoint survey conducted approximately 3 months into the intervention. Data collection was executed by the local Civil Society Organization based in Bandung with the help of peer leaders placed within the key population network (called Focus Muda) in Indonesia. A structured closed-ended questionnaire with special provisions to record open-ended responses wherever required was used in the data collection process. Consent and pre-post assessments were carried out in the local language (Bahasa). All procedures were approved by the Ethical Review Board at the Catholic University of Atmajaya in Jakarta, Indonesia.

Intervention

Development and Design of the mHealth App

For the mobile app to resonate with the Indonesian key populations, in-depth interviews were conducted with a separate

group of individuals who were members of the key populations (N=20). A comprehensive needs assessment to understand the knowledge and experience on HIV-related issues was conducted to determine the feasibility and acceptability of the app in the community. Each in-depth interview lasted for about 45 minutes and was conducted in Bahasa. All interviews were audio-recorded for subsequent transcribing. We conducted content analysis on the interview responses, and the data were coded to extract relevant themes related to the knowledge and experience on HIV-related issues of sexual health, safe sex, and risky behaviors among the key populations in Indonesia, and to identify ideal designing modalities for acceptability of the mobile app in the community. The themes assessed were the availability of smartphones among the respondents and their networks, to understand barriers in accessing services, and to determine the features of the app to inform intervention content, structure, design, and apt elements for a safe and secure mobile app.

Community-Supported Design Process

Following the formative assessment mentioned above, the app interface was designed. Six peer leaders among the 20 who were recruited from Focus Muda—a network of young people working on key population issues across Indonesia—designed the interface of the app in a workshop facilitated by MAMTA-Health Institute for Mother and Child, and the local implementing organization in Bandung. The peer leaders were identified according to preset criteria as community leaders/outreach workers. The main criteria were an association with the network partners, functional literacy and good communication skills, and experience of working with the targeted communities.

The mobile app design was created by peer leaders using a co-design approach [14]. A multistage design process was adopted to finalize the design and features of the app. Stage 1 involved initial assessment of the current situation related to the versions of Android phones under use by the study participants, and to screen resolutions, network connectivity issues, one-time password gateway integration, and related issues. Stage 2 involved having informal discussions with peers on the type of mobile apps regularly used, approximate time spent online, and top-rated features for providing a mobile experience. Stage 3 involved designing the user interface and online avatars for the various features proposed by the peers. Sketching and prototyping the app was then started, and the beta version of the app was tested by different constituencies, including the peers, project implementation team, and technical partners.

Integrated Team of Health Care Providers

A medical doctor with experience in sexually transmitted infections, HIV/AIDS treatment, and management was empaneled with the app. A young primary health care worker with extensive experience in providing HIV counseling and testing services was included within the referral options of the app. Prior approval and consent from the health care providers were obtained before they became involved in the project. The health care providers were given access to the app dashboard (ie, a protected site with a user ID and password) to view the questions asked by the users and respond appropriately within 12 hours. The users were not prescribed medicines online through this app, and the responses were restricted to addressing clarifications and doubts about any of their concerns, including signs and symptoms of HIV/AIDS and other sexually transmitted diseases such as hepatitis C, drug use and rehabilitation, and masturbation.

App Content

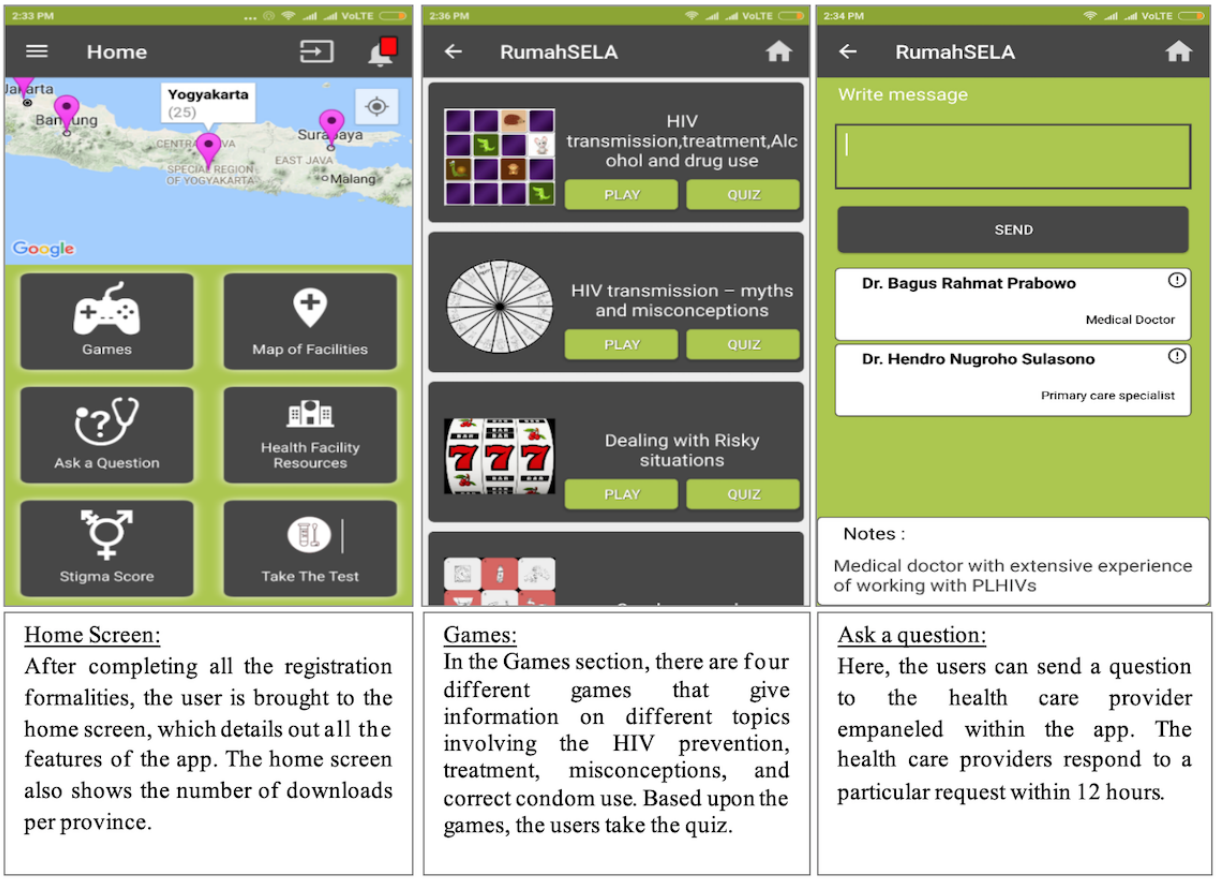
The content included information on the transmission, prevention, and treatment of sexually transmitted infections with a special focus on HIV/AIDS, myths and misconceptions regarding HIV/AIDS, correct condom use, the importance of HIV testing, and access and early linkage to HIV testing services. The app content was focused on these themes as evidence indicates the need to intervene early for increased awareness generation among these populations [15]. The content was provided in two languages, Bahasa and English, and was linked to the internet.

Technical Platform

The final Android-based mobile app was developed by a software development company based in India. The field test of the app was conducted for removal of any bugs and irrelevant content/features before distribution. The mobile app was distributed to each individual's phone through a link, which was secured from the software development company. Registration of the app users, consent, password setting, risk assessment, home screen, ask a question, map of health facilities, request for test intimation, instant in-app "ask a question" facility, stigma information, feedback on the app experience, condom games, quiz, and prize were the key features built into the app. Some screenshots showing the interface of the key features of the app are shown in Figure 1.

The ask-a-question feature helped the users to send private messages to the health care providers empaneled with the app. The health care providers responded to the particular request within 12 hours. Based on the need of every participant, the link to health services was provided to each user through the built-in map of health facilities and the list of health facility resources. This helped the users to easily connect with the health facilities, particularly aimed at improving the linkage and effective utilization of health services. The administration of the app was peer-facilitated, and the participants were reached with a staggered approach. Focus Muda nominated 23 youth peer leaders to administer the mobile app.

Figure 1. Interface of key features of the app.



Measures

Demographics

Demographic variables included age, gender, sexual orientation, relationship status, level of education, current employment, income, living situation, and drug abuse behavior.

HIV/AIDS Knowledge

We assessed the comprehensive knowledge on HIV, adopted and modified from Ankunda and Asiimwe [16], which refers to the knowledge of reducing HIV risk by consistent use of condoms and being faithful to one uninfected sexual partner who has no other partners, and included a rejection of misconceptions of HIV transmission through a mosquito bite and sharing a meal with an HIV-infected person [16]. We considered the knowledge on the first two indicators, and then included the rejection of misconceptions of sharing utensils and clothes with an HIV-infected person and social kissing an HIV-infected person as comprehensive knowledge. This was assessed from baseline findings, which highlighted these myths and misconceptions to be prevalent in the studied key populations. Furthermore, Ankunda and Asiimwe [16] used the knowledge on HIV transmission from an HIV-infected mother to a child as another set of assessment of comprehensive knowledge. During our analysis, we found that knowledge on this indicator was poor in the studied key populations. Therefore, this was adopted as the fifth indicator of comprehensive knowledge on HIV in the current study.

Sexual Behavior

To assess HIV risk behaviors, we asked participants about the number of times they had anal, vaginal, insertive, and receptive sexual intercourse, and whether they had sex with and without a condom with different partners (casual sex with a male or female, transgender male or female) during the last 3 months.

Recency of HIV Testing

We asked the participants “When did you last have an HIV test?” In addition, to assess the effect of the mHealth app on HIV testing, we asked the participants “Did you go for an HIV test after using the app?”

Self-Esteem

We adopted the 10-item Rosenberg Self-Esteem Scale to measure the self-esteem of the participants during the study [17]. The answers to negative items such as “At times, I think I am no good at all” were reverse scored. The scores ranged from 0 to 30 with higher scores indicating higher self-esteem. Cronbach α was .81 and .72 during the pre-post assessments, indicating strong reliability.

Mobile App Acceptability

We also assessed (1) the feasibility of using the mHealth app to receive information on sexual health and HIV (eg, “On what aspects did you receive knowledge through the app?”); (2) perceived impact of using the mHealth app (eg, “Has the app made any impact on your knowledge or behavior? If yes, how has the app made an impact?”); (3) quality of the medico-client

relationship (eg, “Did you send a chat message through the app to get health information? If yes, what kind of health and other information did you seek? Did you feel the response was satisfactory?” [with response options always, sometimes, never] “Did you send a request for booking an appointment for an HIV test with the doctor through the app?”); and (4) general perceptions about the features of the app (eg, “Which feature of the app was most useful in having an impact?” “How would you rate the ease of use of the app?”).

Pre-Post Assessment

Descriptive statistics were computed for demographic characteristics. We applied the McNemar test to assess the possible shift from the baseline to the endpoint for all outcomes. The shift is presented as a percentage point change. Statistical significance was considered at $P \leq .05$. All analyses were carried out using SPSS 22 (IBM Corp, Armonk, NY, USA).

Results

Participation and Intervention Completion Rates

A total of 200 unique users (50 MSM, 50 transgender women, and 100 PWUD) of the app were identified based on the eligibility criteria. Thirty-two participants from the MSM and transgender groups, and 30 from the PWUD group were lost to follow up. Therefore, the follow-up response rate was 98% in MSM and transgender women and was 70% in the PWUD. More than 90% of the dropouts among the PWUD group were injecting drug users.

Sociodemographic Profile

Table 1 summarizes the sociodemographic characteristics of the participants. The mean age was in the low to mid-20s for the three groups. The majority of the PWUD were males. The majority of the participants reported being single at the time of the study. More than half of the participants were educated up to a high school level. More than half of the MSM were working in private sectors, whereas about half of the transgender women were engaged in sex work, followed by singing and dancing. Nearly one-quarter of the PWUD were involved in sex work and over a quarter of the PWUD were working in the private sector. Nearly half of the MSM and about three-quarters of the PWUD reported living with family, whereas the majority of the transgender women lived alone.

The majority of the PWUD reported ever injecting drugs, with the majority indicating that they had injected drugs in the last month and have a history of sharing a needle.

Table 1. Sociodemographic characteristics of the participants (N=168).

Characteristics	Men who have sex with men (n=49)	Transgender women (n=49)	People who use drugs (n=70)
Age (years), mean (SD)	22.6 (3.3)	25.6 (3.0)	23.7 (4.0)
Sex, n (%)			
Male	49 (100)	0	56 (80)
Female	0 (0)	0	14 (20)
Transgender	0 (0)	49 (100)	0
Marital status, n (%)			
Single	49 (100)	45 (92)	56 (80)
Partnered	0 (0)	4 (8)	14 (20)
Educational level, n (%)			
No education	0 (0)	1 (2)	0 (0)
Elementary (grades 1-6)	0 (0)	1 (2)	5 (7)
Middle school (grades 7-9)	1 (2)	14 (29)	13 (19)
High school (grades 10-12)	34 (69)	31 (63)	47 (67)
Technical college/university	14 (29)	2 (4)	5 (7)
Employment, n (%)			
Private job	34 (68)	9 (18)	18 (26)
Government job	2 (4)	0 (0)	0 (0)
Self-employed	3 (6)	6 (12)	13(19)
Singing/Dancing	0 (0)	12 (24)	2 (3)
Sex work	1 (2)	26 (52)	20 (29)
Unemployed	10 (20)	7 (14)	9 (1)
Living situation, n (%)			
Living with parents/family	25 (51)	7 (14)	52 (74)
Living with male partner	6 (12)	2 (4)	5 (7)
Living alone	17 (35)	39 (80)	9 (13)
Living with friends/others	1 (2)	1 (2)	4 (6)
Drug use, n (%)			
Ever taken drugs	1 (2)	5 (10)	70 (100)
Ever injected drugs	0 (0)	2 (4)	58 (83)
Injected drugs in prior 1 month	0 (0)	1 (2)	48 (83)[a]
History of needle exchange	0 (0)	0 (0)	41 (71)[a]

[a]Proportion calculated among those who ever injected drugs.

Correct Knowledge on HIV

Modes of HIV Transmission

Table 2 shows the shift in correct knowledge about the modes of transmission of HIV/AIDS from baseline to the endpoint of the study among the three key populations. The majority of the participants were aware of sex with multiple partners, unprotected sex with an HIV-infected person, blood transfusion from an HIV-infected person, and sharing an HIV-infected needle/syringe as the modes of transmission at baseline. Use of the mobile app resulted in a significant increase in the knowledge about HIV transmission from a mother to child with a percentage point change of 16.4%, 36.7%, and 27.1% among the MSM, transgender women, and PWUD, respectively, from baseline to the endpoint.

Table 2. Shift in correct knowledge about HIV/AIDS from baseline to the study endpoint among the key populations at risk for HIV in Indonesia.

Characteristic	Men who have sex with men (n=49)				Transgender women (n=49)				People who use drugs (n=70)			
	Baseline, n (%)	Endpoint, n (%)	PP[a] change	*P* value	Baseline, n (%)	Endpoint, n (%)	PP change	*P* value	Baseline, n (%)	Endpoint, n (%)	PP change	*P* value
Modes of transmission												
Having sex with multiple partners	47 (96)	49 (100)	4	.16	45 (92)	49 (100)	2	.56	66 (94)	67 (96)	1	>.99
Having unprotected sex with an HIV-infected person	49 (100)	49 (100)	0	.32	49 (100)	49 (100)	0	.08	69 (99)	70 (100)	1	>.99
Having blood transfusion from an HIV-infected person	49 (100)	49 (100)	0	>.99	46 (96)	49 (100)	4	.50	70 (100)	70 (100)	0	.50
Sharing HIV-infected syringes or needles	49 (100)	49 (100)	0	>.99	42 (86)	46 (94)	8	.22	68 (97)	70 (100)	3	.50
From HIV-infected mother to her baby	40 (82)	48 (98)	16	.004	26 (53)	44 (90)	37	<.001	38 (54)	57 (81)	27	.003
Myths and misconceptions about HIV transmission												
HIV can be contracted through mosquito bites	46 (94)	45 (92)	–2	>.99	43 (88)	40 (82)	–6	.51	63 (90)	67 (96)	6	.13
HIV can be contracted by sharing utensils and clothes with an HIV-infected person	44 (90)	45 (92)	2	>.99	45 (92)	49 (100)	8	.13	56 (80.0)	66 (94)	14	.01
HIV can be contracted by living and sharing a meal with an HIV-infected person	36 (92)	47 (96)	4	.63	27 (90)	48 (98)	8	.50	67 (96)	68 (97)	1	>.99
HIV can be contracted by kissing someone who is infected	39 (80)	42 (86)	6	.45	37 (76)	45 (92)	16	.008	56 (80)	63 (90)	10	.04
HIV can be contracted by using public toilets/bathrooms	46 (94)	47 (96)	2	>.99	46 (94)	47 (96)	2	>.99	63 (90)	68 (97)	7	.13
Prevention of HIV												
Using condoms during sexual contact	45 (92)	45 (92)	0	>.99	44 (90)	47 (96)	6	.45	55 (79)	67 (96)	17	.004
Having sex with only one faithful and uninfected partner	41 (84)	89.8	6	.38	28 (57)	38 (78)	21	.02	46 (66)	55 (79)	13	.03
Taking injections using clean and unused syringes	36 (74)	37 (76)	2	>.99	23 (47)	29 (59)	12	.21	55 (79)	63 (90)	11	.02
HIV treatment												
Heard of ART[b] as a treatment of HIV	14 (40)	31 (78)	38	<.001	11 (27)	37 (84)	57	.002	34 (51)	50 (75)	24	<.001
ART can help reduce viral load	30 (61)	35 (71)	10	.56	34 (69)	35 (71)	2	.32	35 (50)	45 (64)	14	.27

Characteristic	Men who have sex with men (n=49)				Transgender women (n=49)				People who use drugs (n=70)			
	Baseline, n (%)	Endpoint, n (%)	pp[a] change	*P* value	Baseline, n (%)	Endpoint, n (%)	PP change	*P* value	Baseline, n (%)	Endpoint, n (%)	PP change	*P* value
ART can help improve quality of life	22 (63)	29 (73)	10	.32	29 (59)	31 (74)	15	.66	22 (31)	41 (9)	27	<.001

[a]PP: percentage point.

[b]ART: antiretroviral therapy.

Myths and Misconceptions

The increase in the awareness that HIV does not spread by sharing utensils and clothes was most significant among PWUD with a percentage point change of 14%. Over three-quarters of transgender women and PWUD reported that a person can contract HIV by kissing someone who is infected before the intervention. There was a significant change in the level of awareness with regard to this myth postintervention, with a percentage point change of 16% and 10% for transgender women and PWUD respectively reporting the correct information (Table 2).

HIV Prevention

Over three-quarters of the PWUD correctly reported that condom use can prevent HIV transmission before the intervention, and this knowledge further increased after the intervention with a percentage point change of 17% (Table 2).

HIV Treatment

Knowledge on ART as the basic line of treatment for HIV significantly increased among all groups postintervention by 38%, 57%, and 24% among MSM, transgender women, and PWUD, respectively (Table 2).

Comprehensive HIV Knowledge

There were significant shifts in comprehensive knowledge on HIV from 20% (10/49) to 60% (29/49), 22% (11/49) to 57% (28/49), 49% (34/70) to 74% (52/70) among MSM (percentage point change of 40%, *P*=.004), transgender women (percentage point change of 35%, *P*<.001), and PWUD (percentage point change of 25%, *P*<.001), respectively, representing a total change from 29% to 63% for all key populations.

HIV-Related Risk Behaviors

There was a reduction in the number of individuals who did not use a condom in their last sexual intercourse encounter postintervention in all three groups (with a maximum shift in the PWUD). Among PWUD who abused drugs in the form of injection, the number of individuals who reported injecting drugs with a used needle or syringe in the last month reduced by over 50% after the intervention (Table 3).

Table 3. High-risk behaviors among key populations in the baseline and endpoint surveys.

Indicators	Men having sex with men (n=49)				Transgender women (n=49)				People who use drugs (n=70)			
	Baseline, n (%)	Endpoint, n (%)	pp[a] change	*P* value	Baseline, n (%)	Endpoint, n (%)	PP change	*P* value	Baseline, n (%)	Endpoint, n (%)	PP change	*P* value
Not using condoms at the last sexual intercourse	11 (22)	9 (19)	–3	.45	9 (18)	6 (12)	–6	.25	15 (21)	7 (10)	–11	<.001
Injecting with a used needle or syringe in the past 3 months	0	0	0	N/A[b]	1 (2)[c]	0	–2	N/A	45 (78)	15 (26)	–52	<.001

[a]PP: point percentage.

[b]N/A: not applicable.

[c]N=1 among transgender women who abused drugs in the form of injection; therefore, no statistics were computed.

Access to HIV Testing Services

There was a significant increase in the uptake of HIV testing after using the mHealth app from 79% (133/168) to 90% (151/168) before and after the intervention with a percentage point change of 11% (*P*<.035) for the three groups combined. The shift was the highest in the PWUD group, from 65% (46/70) before the intervention to 82% (57/70) after the intervention with a percentage point change of 17% (*P*<.001), followed by MSM (from 88% to 94%) and transgender women (96% to 98%).

Overall, 34% of the study participants reported that they went for an HIV test after using the app, 21% (35/168) of whom had made the appointment through the app. Almost half (24/49, 49%) of the transgender women, and 31% (15/49) and 26% (18/70) of the MSM and PWUD reported that they went for an HIV test after using the app. Of those who reported going for an HIV test after using the app, 27% (7/24) MSM, 25% (4/15) transgender women, and 11% (2/18) PWUD had made the appointment through the app.

Self-Esteem

There was a positive shift in self-esteem from baseline to the endpoint in all groups with a significant shift among the transgender women (Table 4).

Table 4. Self-esteem scores among the key populations in the baseline and the endpoint surveys.

Population	Baseline, mean (SD)	Endpoint, mean (SD)	*t* value	*P* value
Men who have sex with men (n=49)	26.44 (3.66)	27.06 (3.41)	0.873	.39
Transgender women (n=49)	26.54 (2.83)	27.84 (2.38)	–2.468	.02
People who use drugs (n=70)	24.04 (5.45)	25.04 (3.58)	–1.324	.19

Acceptability of the App

The majority of the participants (36/39, 74% MSM; 41/49, 84% transgender women; and 47/70, 67% PWUD) stated that they liked the game and health facility map features of the app. Over one-quarter (13/49, 26%) MSM and transgender women each, and 16% (11/70) of the PWUD asked a question through the app to get health information. The most common type of information sought was knowledge about HIV. The majority of the participants (40/49, 82% MSM; 41/49, 84% transgender women; 59/70, 84% PWUD) felt that the app provided knowledge on HIV with respect to risk, safer sexual practices, and prevention and treatment of HIV. Nearly a quarter of the participants (11/49, 22% MSM; 10/49, 20% transgender women; 16/70, 23% PWUD) said that they became more confident in discussing issues on sexuality after using the app. Nearly half of the participants (28/49, 57% MSM; 15/49, 31% transgender women; 41/70, 59% PWUD) found the app neither difficult nor easy to use, while 33% (16/49) MSM and transgender women each, and 20% (14/70) PWUD found the app easy to use. Further, only 23% (39/168) of the total participants had used any other mobile app before RUMAH SELA.

Discussion

Principal Findings

This study tested the potential use of an mHealth app (RUMAH SELA) as a self-learning tool for HIV prevention, which was developed and designed internally, and customized for key populations (MSM, transgender women, and PWUD) in Indonesia with the support of peer leaders. This intervention shows potential promise for enhancing the spectrum of HIV knowledge, reduction in risky sexual behaviors, uptake of services, and improving self-esteem among these vulnerable and underserved groups.

The comparison of outcomes before and after the intervention showed marked improvement in HIV-related knowledge, testing, and behavioral outcomes (eg, condom use and not injecting drugs using syringes/shared needles among PWUD). There was a potential uptake of services after the intervention. This study showed that 34% of the participants took an HIV test after using the app. A significant reduction was observed in the number of individuals who did not use condoms in their last sexual intercourse encounter postintervention in all three groups. PWUD showed a marked reduction in injecting drugs using a shared needle after using the mHealth app. Similar intervention strategies such as internet-based, text messaging, and smartphone apps were also found to be successful in reducing substance abuse in studies conducted in upper-middle and high-income countries such as Romania, the United States, Germany, and Switzerland in 2017 [18,19].

We found that certain myths regarding the transmission of HIV through a mosquito bite, social kissing, and sharing common utensils/clothes/bathroom/meal with an infected person were prevalent among the key populations, with the spread of HIV through social kissing as the most prevalent myth believed. This mHealth intervention could successfully dispel these myths.

There was a 3%, 6%, and 11% reduction among MSM, transgender women, and PWUD, respectively, who did not use a condom in their last intercourse after the intervention, indicating the potential value of the mHealth app in enhancing safer sexual practices. The study findings are in line with other studies that have discerned the impact of mHealth interventions in improving the sexual and behavioral health of vulnerable and underserved populations [12,18,20,21].

Although assessment of mental health was not within the scope of the present intervention, we tried to assess the effect of the app on self-esteem of the participants. It is noteworthy that the use of the mHealth app in this intervention could help to raise the self-esteem of the participants. Self-esteem is a precursor to both physical and mental health. Further, it acts as a protective factor contributing to positive social behavior [22]. Improving the self-esteem of an underserved population may have a benefit by acting as a catalytic agent for behavior change and consequently contributing to increased seeking for HIV services, prevention, and care.

In addition to behavioral interventions, the combination of access to HIV services along with behavioral change is an important strategy for HIV prevention and care. Most existing mobile-based interventions are contextualized particularly to improve the medical adherence to ART for people living with HIV [13,23-25]. Few qualitative studies have emphasized the use of mHealth for promoting the uptake of HIV testing [16,26]. This study highlights the utility of an mHealth app designed on the principles of community engagement in not only creating awareness about safer sex and increasing knowledge about HIV prevention, transmission, and treatment, but also in improving the uptake of HIV services among young key populations. A change was also found in terms of users having learned about the location of health service providers and the self-assessment of risk for HIV. The features of the app that were deemed to have the most positive impact were the map of health service facilities, games, and the ability to share your story, aside from acknowledging the user-friendliness of the app as key aspects that highlighted app acceptability by the target population.

Strengths and Limitations

One of the strengths of this study is that it is a peer-customized mHealth app useful for HIV interventions based on a self-learning principle with minimal manual involvement. However, the findings of this study need to be considered in the context of certain limitations. The small sample size was one of the major limitations of this study; we assume that a larger sample could have provided more statistically meaningful inferences. The retention rate was 98% among MSM and transgender women but was only 70% among PWUD.

This study did not include any control group, which limits the validation of the intervention's efficacy in changing the sexual and behavioral health outcomes. Further, based on the formative results, Facebook, Twitter, Instagram, Lollipop, Grinder, and Hornet were listed as apps/social media platforms that the participants had already used. These apps could be other possible potential sources of information for these populations apart from the RUMAH SELA app. However, the longitudinal nature of the study did provide an opportunity to test the efficacy of the intervention. Another limitation of the study was that it was a short-term intervention with an endpoint assessment conducted after 3 months. An intervention conducted for a longer duration with periodic assessments may have yielded more informative results. Since the app was designed and customized for the key populations of Indonesia, its utility may not be generalized for a general population. Lastly, we could not assess the impact of the sociodemographic parameters on the intervention for the differential outcomes related to the utility of the app across the groups due to the small sample size and thus insufficient statistical power for inference. Nevertheless, the app was found to be fairly well accepted by users. Moreover, only less than a quarter of the study participants had used any other mobile app before RUMAH SELA, which supports to a certain degree that the shift in knowledge and health-seeking behaviors among the study participants was directly due to exposure to RUMAH SELA.

Conclusions

The RUMAH SELA mHealth app opens a window for enhancing knowledge on HIV and related behavioral outcomes among key populations with minimal external support and involvement, thereby offering a cost-effective and self-administered intervention package. This calls for larger intervention studies/trials, small-scale efforts, and pilot studies that must begin to scale with a rigorous design for future evaluation. These efforts will strengthen and provide evidence for program leaders, policymakers, and funders to develop influential population-based mHealth intervention strategies to combat HIV.

Acknowledgments

We acknowledge the support of Rumah Chemara Indonesia for hosting the program and research in Indonesia. Further, we thank the peer leaders from Indonesia who participated in the design and dissemination of the app in the community. We extend special thanks to Mr. Sumit Basu who took the lead in conducting the SOGI (Sexual Orientation and Gender Identity) training for peer leaders in Indonesia. We also would like to extend thanks to Focus Muda for designating peer leaders for the study. We are grateful to the group of doctors who agreed to be empaneled with the app, especially to be available to respond to queries by the participants. Last but not the least, we extend thanks to all of the participants of the study. The funding for the program and research was provided by the Netherlands Ministry of Foreign Trade and Development Cooperation.

Conflicts of Interest

None declared.

References

1. Pendse R, Gupta S, Yu D, Sarkar S. HIV/AIDS in the South-East Asia region: progress and challenges. J Virus Erad 2016 Nov 28;2(Suppl 4):1-6 [FREE Full text] [Medline: 28303199]
2. WHO Regional Office for South-East Asia. HIV/AIDS in the South-East Asia Region: progress report 2011. World Health Organization. 2012. URL: https://apps.who.int/iris/handle/10665/205306 [accessed 2018-07-17]
3. Poudel K, Jimba M. HIV care continuum for key populations in Indonesia. Lancet HIV 2018 Oct;5(10):e539-e540. [doi: 10.1016/S2352-3018(18)30169-3] [Medline: 30143456]
4. Ooms G, Kruja K. The integration of the global HIV/AIDS response into universal health coverage: desirable, perhaps possible, but far from easy. Global Health 2019 Jun 18;15(1):41 [FREE Full text] [doi: 10.1186/s12992-019-0487-5] [Medline: 31215446]
5. Januraga PP, Reekie J, Mulyani T, Lestari BW, Iskandar S, Wisaksana R, et al. The cascade of HIV care among key populations in Indonesia: a prospective cohort study. Lancet HIV 2018 Oct;5(10):e560-e568. [doi: 10.1016/S2352-3018(18)30148-6] [Medline: 30143455]
6. Indonesia country profile at a glance: National AIDS Commission of Indonesia, 2011. Evidence to action. 2011. URL: http://www.aidsdatahub.org/Country-Profiles/Indonesia [accessed 2018-07-17]
7. Hampshire K, Porter G, Owusu SA, Mariwah S, Abane A, Robson E, et al. Informal m-health: How are young people using mobile phones to bridge healthcare gaps in Sub-Saharan Africa? Soc Sci Med 2015 Oct;142:90-99 [FREE Full text] [doi: 10.1016/j.socscimed.2015.07.033] [Medline: 26298645]

8. Lee SS, Tam DKP, Mak DWL, Wong KH. Use of the Internet for sex partnership in men who have sex with men before HIV infection. Public Health 2011 Jul;125(7):433-435. [doi: 10.1016/j.puhe.2011.03.010] [Medline: 21733532]
9. Justumus P, Colby D, Mai Doan Anh T, Balestre E, Becquet R, Orne-Gliemann J. Willingness to use the Internet to seek information on HIV prevention and care among men who have sex with men in Ho Chi Minh City, Vietnam. PLoS One 2013;8(8):e71471 [FREE Full text] [doi: 10.1371/journal.pone.0071471] [Medline: 23977048]
10. Anand T, Nitpolprasert C, Kerr SJ, Apornpong T, Ananworanich J, Phanuphak P, et al. Implementation of an online HIV prevention and treatment cascade in Thai men who have sex with men and transgender women using Adam's Love Electronic Health Record system. J Virus Erad 2017 Jan 01;3(1):15-23 [FREE Full text] [Medline: 28275454]
11. The state of mobile internet connectivity report 2019. Mobile for Development. 2019. URL: https://www.gsma.com/mobilefordevelopment/resources/the-state-of-mobile-internet-connectivity-report-2019/ [accessed 2019-07-20]
12. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
13. Youth in Indonesia. UNFPA Indonesia Monograph Series No. 2. 2014. URL: https://indonesia.unfpa.org/sites/default/files/pub-pdf/BUKU_Monograph_No2_Youth_in_Indonesia_ENG_05_Low-res.pdf [accessed 2018-07-17]
14. Kensing F, Blomberg J. Participatory Design: Issues and Concerns. Comput Support Coop Work 1998 Sep;7(3-4):167-185. [doi: 10.1023/a:1008689307411]
15. Bekker L, Hosek S. HIV and adolescents: focus on young key populations. J Int AIDS Soc 2015 Feb 26;18(2 Suppl 1):20076. [doi: 10.7448/IAS.18.2.20076]
16. Ankunda D, Asiimwe JB. Determinants of comprehensive knowledge of HIV/AIDS among women of the reproductive age (15-49) in Uganda. Int J Community Med Public Health 2017 Sep 22;4(10):3530. [doi: 10.18203/2394-6040.ijcmph20174215]
17. Coetzee J, Buckley J, Otwombe K, Milovanovic M, Gray GE, Jewkes R. Depression and Post Traumatic Stress amongst female sex workers in Soweto, South Africa: A cross sectional, respondent driven sample. PLoS One 2018;13(7):e0196759 [FREE Full text] [doi: 10.1371/journal.pone.0196759] [Medline: 29975685]
18. Lelu iu-Weinberger C, Manu M, Ionescu F, Dogaru B, Kovacs T, Dorobăn escu C, et al. An mHealth Intervention to Improve Young Gay and Bisexual Men's Sexual, Behavioral, and Mental Health in a Structurally Stigmatizing National Context. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e183 [FREE Full text] [doi: 10.2196/mhealth.9283] [Medline: 30429117]
19. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A Systematic Review of the mHealth Interventions to Prevent Alcohol and Substance Abuse. J Health Commun 2017 May;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
20. Chiang N, Guo M, Amico KR, Atkins L, Lester RT. Interactive Two-Way mHealth Interventions for Improving Medication Adherence: An Evaluation Using The Behaviour Change Wheel Framework. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e87 [FREE Full text] [doi: 10.2196/mhealth.9187] [Medline: 29650504]
21. Tucker JD, Tso LS, Hall B, Ma Q, Beanland R, Best J, et al. Enhancing Public Health HIV Interventions: A Qualitative Meta-Synthesis and Systematic Review of Studies to Improve Linkage to Care, Adherence, and Retention. EBioMedicine 2017 Mar;17:163-171 [FREE Full text] [doi: 10.1016/j.ebiom.2017.01.036] [Medline: 28161401]
22. Mann M, Hosman CMH, Schaalma HP, de Vries NK. Self-esteem in a broad-spectrum approach for mental health promotion. Health Educ Res 2004 Aug;19(4):357-372. [doi: 10.1093/her/cyg041] [Medline: 15199011]
23. Whiteley L, Brown L, Lally M, Heck N, van den Berg JJ. A Mobile Gaming Intervention to Increase Adherence to Antiretroviral Treatment for Youth Living With HIV: Development Guided by the Information, Motivation, and Behavioral Skills Model. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e96 [FREE Full text] [doi: 10.2196/mhealth.8155] [Medline: 29685863]
24. Bisallah CI, Rampal L, Lye M, Mohd Sidik S, Ibrahim N, Iliyasu Z, et al. Effectiveness of health education intervention in improving knowledge, attitude, and practices regarding Tuberculosis among HIV patients in General Hospital Minna, Nigeria - A randomized control trial. PLoS One 2018;13(2):e0192276 [FREE Full text] [doi: 10.1371/journal.pone.0192276] [Medline: 29470530]
25. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth Interventions To Support Self-Management In HIV: A Systematic Review. Open AIDS J 2017;11:119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
26. Evans C, Turner K, Suggs LS, Occa A, Juma A, Blake H. Developing a mHealth intervention to promote uptake of HIV testing among African communities in the UK: a qualitative study. BMC Public Health 2016 Jul 28;16:656 [FREE Full text] [doi: 10.1186/s12889-016-3278-4] [Medline: 27465586]

Abbreviations

ART: antiretroviral therapy
mHealth: mobile health
MSM: men who have sex with men

PWUD: people who use drugs

Edited by G Eysenbach; submitted 31.12.19; peer-reviewed by P Dhall, M Ghule; comments to author 10.02.20; revised version received 31.05.20; accepted 25.06.20; published 08.09.20.

Please cite as:
Garg PR, Uppal L, Mehra S, Mehra D
Mobile Health App for Self-Learning on HIV Prevention Knowledge and Services Among a Young Indonesian Key Population: Cohort Study
JMIR Mhealth Uhealth 2020;8(9):e17646
URL: https://mhealth.jmir.org/2020/9/e17646
doi:10.2196/17646
PMID:32896831

Original Paper

Validity and Usability of a Smartphone Image-Based Dietary Assessment App Compared to 3-Day Food Diaries in Assessing Dietary Intake Among Canadian Adults: Randomized Controlled Trial

Yuwei Ji[1], MSc, RD; Hugues Plourde[1], PhD, RD; Valerie Bouzo[1], BSc; Robert D Kilgour[2,3], PhD; Tamara R Cohen[3,4], PhD, RD

[1]School of Human Nutrition, McGill University, Montreal, QC, Canada

[2]Department of Health, Kinesiology, and Applied Physiology, Concordia University, Montreal, QC, Canada

[3]PERFORM Centre, Concordia University, Montreal, QC, Canada

[4]Faculty of Land and Food Systems, Food, Nutrition and Health, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Tamara R Cohen, PhD, RD
Faculty of Land and Food Systems
Food, Nutrition and Health
University of British Columbia
218-2205 East Mall
Vancouver, BC, V6T 1Z4
Canada
Phone: 1 604 827 0362
Email: tamara.cohen@ubc.ca

Abstract

Background: Accurate dietary assessment is needed in studies that include analysis of nutritional intake. Image-based dietary assessment apps have gained in popularity for assessing diet, which may ease researcher and participant burden compared to traditional pen-to-paper methods. However, few studies report the validity of these apps for use in research. Keenoa is a smartphone image-based dietary assessment app that recognizes and identifies food items using artificial intelligence and permits real-time editing of food journals.

Objective: This study aimed to assess the relative validity of an image-based dietary assessment app — Keenoa — against a 3-day food diary (3DFD) and to test its usability in a sample of healthy Canadian adults.

Methods: We recruited 102 participants to complete two 3-day food records. For 2 weeks, on 2 non-consecutive days and 1 weekend day, in random order, participants completed a traditional pen-to-paper 3DFD and the Keenoa app. At the end of the study, participants completed the System Usability Scale. The nutrient analyses of the 3DFD and Keenoa data before (Keenoa-participant) and after they were reviewed by dietitians (Keenoa-dietitian) were analyzed using analysis of variance. Multiple tests, including the Pearson coefficient, cross-classification, kappa score, % difference, paired t test, and Bland-Altman test, were performed to analyze the validity of Keenoa (Keenoa-dietitian).

Results: The study was completed by 72 subjects. Most variables were significantly different between Keenoa-participant and Keenoa-dietitian (P<.05) except for energy, protein, carbohydrates, fiber, vitamin B1, vitamin B12, vitamin C, vitamin D, and potassium. Significant differences in total energy, protein, carbohydrates, % fat, saturated fatty acids, iron, and potassium were found between the 3DFD and Keenoa-dietitian data (P<.05). The Pearson correlation coefficients between the Keenoa-dietitian and 3DFD ranged from .04 to .51. Differences between the mean intakes assessed by the 3DFD and Keenoa-dietitian were within 10% except for vitamin D (misclassification rate=33.8%). The majority of nutrients were within an acceptable range of agreement in the Bland-Altman analysis; no agreements were seen for total energy, protein, carbohydrates, fat (%), saturated fatty acids, iron, potassium, and sodium (P<.05). According to the System Usability Scale, 34.2% of the participants preferred using Keenoa, while 9.6% preferred the 3DFD.

Conclusions: The Keenoa app provides acceptable relative validity for some nutrients compared to the 3DFD. However, the average intake of some nutrients, including energy, protein, carbohydrates, % fat, saturated fatty acids, and iron, differed from

the average obtained using the 3DFD. These findings highlight the importance of verifying data entries of participants before proceeding with nutrient analysis. Overall, Keenoa showed better validity at the group level than the individual level, suggesting it can be used when focusing on the dietary intake of the general population. Further research is recommended with larger sample sizes and objective dietary assessment approaches.

(JMIR Mhealth Uhealth 2020;8(9):e16953) doi:10.2196/16953

KEYWORDS

mobile food record; validity; image-based dietary assessment; healthy adults; 3-day food diary; diet; application; nutrition; mHealth; Canada

Introduction

Assessment of dietary intake, in particular habitual dietary intake, remains a major challenge for researchers [1]. Limitations of dietary assessment have been well documented and vary by the method chosen. Research typically relies on traditional pen-and-paper methods to assess dietary intake: the 3-day food diary (3DFD), 24-h recall, or food frequency questionnaire (FFQ) [2]. However, researchers face many challenges when deciding which method is best [2]. Issues of participant burden, motivation and willingness to accurately report diet, and participant literacy and memory should be considered. Moreover, time to enter and analyze diet data, and therefore the availability of resources to correctly analyze dietary recalls, should be accounted for before commencing a study.

All methods of dietary assessment have their limitations [3]. For instance, a single 24-h recall only reflects the food consumed from a single random day and may be less representative of an estimated individual's intake. However, two or more 24-h recalls or food records are needed to estimate usual dietary intake distributions [4]. A limitation of food records is that they can lead to reactivity bias [5]. FFQs may lead to over reporting of average dietary intakes and, similar to 24-h recalls, rely on the participant's ability to correctly recall portion sizes and frequencies [5]. Further, the FFQ can be lengthy; therefore, time and ability to remain focused can be demanding for participants [6]. For these reasons, new methods of dietary assessment are needed to benefit participants and researchers.

With the development of technology, image-based dietary records have increased in popularity. There are different technology-based methods of tracking diet, including taking pictures [7] or using digital food databases [2,7,8]. Relevant to this study, mobile photo meal apps engage the user to take a picture of their food item(s) using their mobile device before it is consumed and require the user to input the quantity and other pertinent details, as needed. Although mobile phone–based methods still rely on user input, an advantage of this method is that food entries are time-stamped, which identifies when food items were consumed [9]. The use of dietary assessment applications, such as those that use a mobile phone app, has shown an increase in participant satisfaction and preference compared with conventional methods (eg, 24-h recall, written food diary) [2]. Moreover, using mobile devices has the potential to reduce costs and diet data entry errors by researchers [7,10]. However, assessment of the validity of these apps against traditional methods of dietary assessment is needed to ensure participant satisfaction and therefore compliance to recording intakes.

Keenoa (Montreal, Quebec, Canada) is a smartphone imaged–based dietary assessment app that recognizes and identifies food items using artificial intelligence and permits editing of food journals in real time. Unique to other apps, Keenoa is accessed only by registered dietitians licensed to practice in Canada with the idea that dietitians are trained to identify food items that were missed or misidentified by the user. Therefore, the advantage of Keenoa is that dietitians can adjust the food items to generate accurate nutrient profiles of an individual's dietary intake. Currently, the app is being used by practicing dietitians in Canada. From a researcher's perspective, using Keenoa to assess dietary intakes would reduce systematic errors associated with data entry [5]. It is currently unknown if the app is appropriate for use in research.

This study aimed to assess the relative validity of Keenoa against the 3DFD and to test its usability in a sample of healthy Canadian adults.

Methods

Participants and Recruitment

From February to April 2019, we recruited 102 participants at the PERFORM Centre (Concordia University, Montreal, Quebec, Canada). Inclusion criteria included adults (>18 years of age) who owned a smartphone (Android or Apple) and would be able to download the Keenoa app without assistance. Exclusion criteria included individuals with a previously diagnosed disease affecting their dietary intake (ie, Type 1 or Type 2 diabetes, renal disease, inflammatory or immunity disorders); currently following a diet or weight-loss regime; who, as an adult, suffered from, had a history of, or was being treated for an eating disorder; or who have completed or are in the process of completing a nutrition or related degree (eg, a dietetics major). Finally, individuals who were not French or English speaking and who could not understand written English were also excluded. Compensation for completing the study included a detailed dietary assessment by a registered dietitian that was emailed to them. This study was approved by the University Human Research Ethics Committee of Concordia University, and written consent was obtained from all participants during their first visit at the PERFORM Centre.

Procedure

Each participant met with a trained researcher (VB, TC) at the PERFORM Centre. Participants completed a brief

sociodemographic questionnaire with questions related to total family income, ethnicity, highest level of education, age, and sex. Body weight and height were measured using a balance scale and a wall-mounted height rod, respectively. BMI was calculated using the measured weight and height (kg/m^2).

Using a computer-generated list, participants were randomly assigned to start with either the Keenoa application or 3DFD for the first week and then switch methods of dietary assessment for the second week. Both 3DFD and Keenoa recordings needed to include 2 non-consecutive weekdays and 1 weekend day. Participants were highly encouraged to maintain their typical eating habits throughout the study.

Participants were provided a calendar to help them keep track of the specific days they should record their diet using the respective methods. A trained researcher also reviewed how to estimate portion sizes using the Dietitian's of Canada Handy Guide to Servings Sizes [11].

3-Day Food Diary (3DFD)

All participants were provided a hard-copy, single-page printout of the standard 3DFD to record their diet. On the 3DFD, there were prompts for participants to record and estimate the details of all meals and snacks they consumed including portion size, cooking methods, and any add-ons (eg, cream, oil, butter, jam). If they had meals in restaurants, participants were told that the restaurant name and dishes needed to be recorded in detail as well. Participants were given the option to either mail the 3DFD with a prestamped posted envelope directed to the corresponding author or scan and email their 3DFD to the study email. All 3DFD were coded for confidentially; those emailed were immediately printed, and the email was deleted to maintain confidentially.

Keenoa App

Instructions on how to use the app were provided to the participant, and online tutorials were available from the company. In the case of this study, the researcher used the participant's email to send an invitation to download the app on their smartphone, which, once downloaded, automatically connected the user to the study dietitian. Participants were asked to take pictures of the food items before they were consumed using their smartphone. Items could range from a single item (eg, an apple, French fries, a cup of coffee) to composite items (eg, a serving of lasagna, bowl of soup, or slice of pizza). If the app recognized the food item(s), it would display options for the users to choose from to correctly identify the food(s). Otherwise, participants could search and record the foods manually from a database that was linked to the Canadian Nutrient File (2015) [12]. The Canadian Nutrient File is a bilingual (English and French) food composition database that is managed by the Government of Canada that includes foods available only on the Canadian market.

Once the food item was identified, the app then prompted participants to estimate and enter the serving size. Using visual aids (ie, a tennis ball, picture of a measuring cup), the participant scrolled to the correct unit (ie, unit count, volume, or weight) and identified a number corresponding to the unit. If a participant did not consume the entire food item, they had the option to record the information in text in the app. If a participant forgot to take a picture prior to consuming a food item or meal, they also had the option to manually enter foods and text in the app.

Once a meal was complete, the user finalized the day, at which point the image and corresponding information were immediately uploaded to the research dietitians' private page. Food items' nutrient values were automatically computed using nutrient information from the Canadian Nutrient File (2015) database. Once participants completed their 3 days using Keenoa, the nutrient analyses were exported to Excel and coded to maintain confidentiality.

Nutrient Analyses

All the 3DFD data were reviewed and recorded in the Food Processor software (ESHA Research version 11.1, Oak Brook, IL) by a trained researcher. The 3-day average nutrient intake content for each subject was computed by the software for total energy (kcal), fat (g), protein (g), carbohydrate (g), saturated fatty acids (g), cholesterol (g), dietary fiber (g), and micronutrients, including vitamin A (µg), vitamin B1 (mg), vitamin B2 (mg), vitamin B12 (µg), vitamin C (mg), vitamin D (µg), calcium (mg), iron (mg), magnesium (mg), phosphorus (mg), potassium (mg), and sodium (mg). All foods were entered into the software using the Canadian database, which uses the Canadian Nutrient File (2015) dataset [12]. Data were exported to Excel for statistical analysis.

The food records from Keenoa were also exported exactly as they were entered by the participant (Keenoa-participant). A research dietitian then reviewed each food record for missing items or misrecorded portion sizes as per the images (Keenoa-dietitian) and exported these data for analysis to Excel. The nutrition assessment and recommendations were sent to the participants by the research dietitian after the reports were corrected and reviewed by the dietitian. Similar to the Food Processor software (ESHA) and as mentioned, Keenoa also uses the Canadian Nutrient File (2015) database for diet analysis; the same nutrients were exported as for the 3DFD analysis.

Exit Survey

After completing both methods of dietary assessment, the participants were sent a link to complete an online survey. This English survey included the System Usability Scale (SUS) questionnaire [13] with 3 additional questions related to using the 3DFD method and Keenoa app. The SUS has been used in previous research to examine the user's perspective of a mobile app [14,15]. This questionnaire includes 10 items and uses a 5-item rating scale. Specifically, this questionnaire surveys participants on different aspects of the Keenoa app (eg, adoption and complexity).

Statistical Analyses

The final analyses were restricted to those who completed both methods of diet recall (n=72), thereby excluding participants who withdrew (n=30). The characteristics of participants, as well as the results of the survey, are presented as percentages, means, and SD. Chi-square tests and Student's *t* tests were performed to identify the differences in demographic and

socioeconomic characteristics between participants who completed the study and those who dropped out. The mean and SD of the nutrition intake of the Keenoa-participant data, Keenoa-dietitian data, 3DFD data, and percentage of energy present in the macronutrients (protein, carbohydrates, and fat) were calculated.

Repeated analyses of variance (ANOVA) with post hoc tests were used to compare the differences in nutrient consumption among the 3 groups. Percentage differences (% difference) between Keenoa-participant versus Keenoa-dietitian and Keenoa-dietitian versus 3DFD were also calculated.

As the purpose of using the Keenoa app is to analyze participants' or clients' dietary intake after adjustment by registered dietitians, 6 cross-classification analyses were performed (ie, Pearson coefficient, cross-classification, kappa score, % difference, *t* test, Bland-Altman test) to assess the validity of the Keenoa app (Keenoa-dietitian). Weighted Cohen kappa and cross-classification tests were performed to evaluate the interrater agreement between the diet data from Keenoa-dietitian and 3DFD. This was analyzed by calculating the chance of misclassification between the 2 methods (eg, a participant being classified in the first quartile by 3DFD but classified in the fourth quartile by Keenoa or vice versa). Pearson correlations and Bland-Altman tests were also used to test for associations between Keenoa-dietitian and 3DFD. Validity assessments were performed as suggested by Lombard and colleagues [16], which combines the results of the 6 tests mentioned. The result of each test was classified as "good," "acceptable," or "poor," and the total number of poor outcomes was calculated. All analyses were performed in SPSS version 23 (IBM Inc, Armonk, NY) with *P* values <.05 considered statistically significant.

Results

We recruited 102 participants in this study; 6 individuals did not attend the baseline visit, while 21 participants did not complete the study protocol as directed (ie, using one of two methods on 3 non-consecutive days to record diet, did not complete full days of recording food items on Keenoa**,** or did not return their 3DFD as instructed.) Both 3DFD and Keenoa diet reports were completed per the study protocol by 75 participants. Due to outlying diet data that could not be edited by the system, 3 participants were excluded from the analysis. This study reports on women (n=47) and men (n=25) with a combined mean age of 38.5 years and a mean BMI of 27.0 kg/m^2 (SD 4.9 kg/m^2; Table 1). Most participants (80.6%) held a university degree, and 62.5% of the participants had a family income of more than Can $30,000 per year. The sample population was 56.9% Caucasian, 19.4% Asian, and 23.6% other ethnicities, including Arab, African American, and Latin American. Only 2 participants (2.8%) reported having a vegan or vegetarian diet. There was no significant difference in characteristics in those who completed the study and those who withdrew.

Table 1. Participant demographic characteristics (n=72).

Characteristics	Values
Age (years), mean (SD)	38.5 (16.7)
Age (years), n (%)	
18-30	31 (43.1)
31-50	18 (25.0)
51-65	17 (23.6)
≥65	6 (8.3)
BMI (kg/m^2), mean (SD)	27.0 (4.9)
BMI (kg/m^2), n (%)	
<25	24 (33.3)
25 to <30	31 (43.1)
≥30	17 (23.6)
Sex, n (%)	
Female	47 (65.3)
Male	25 (34.7)
Education, n (%)	
College and below	12 (16.7)
University	58 (80.6)
Refused to answer	2 (2.8)
Family income (Can $), n (%)	
<30 000	14 (19.4)
≥30 000	45 (62.5)
Refused to answer	13 (18.1)
Ethnicity, n (%)	
White	41 (56.9)
Asian	14 (19.4)
Other	17 (23.6)
Vegetarian or vegan diet, n (%)	
Yes	2 (2.8)
No	70 (97.2)

The differences between Keenoa-participant, Keenoa-dietitian, and 3DFD data are presented in Table 2. The percentage mean intakes of protein (*P*=.001), carbohydrates (*P*<.001), and fat (*P*<.001) were significantly different between Keenoa-participant and Keenoa-dietitian data as were grams of fat (*P*<.001), saturated fatty acids (*P*<.001), cholesterol (*P*<.001), vitamin A (*P*<.001), vitamin B2 (*P*<.001), magnesium (*P*=.009), phosphorus (*P*<.001), iron (*P*<.001), and sodium (*P*<.001). The majority of nutrients from Keenoa-participant were under-recorded compared with Keenoa-dietitian, excluding % protein, carbohydrates, % carbohydrates, and calcium. Vitamin A showed the highest percentage difference (134.2%, data not shown), and the lowest percentage difference was observed for potassium (0.7%, data not shown).

Table 2. Differences between nutrients from Keenoa-participant, Keenoa-dietitian, and 3-day food diary (3DFD; n=72).

Nutrients	Keenoa-participant, mean (SD)	Keenoa-dietitian, mean (SD)	3DFD, mean (SD)	*P* value
Energy (kcal)	1615.3 (1664.4)	1693.0 (593.2)[a]	2006.3 (540.5)	.000
Protein (g)	65.0 (40.7)	68.8 (24.9)[a]	85.6 (26.0)[b]	.000
% Protein	18.0 (5.4)[c]	16.5 (3.6)	17.4 (4.2)	.001
Carbohydrate (g)	225.5 (372.5)	181.8 (65.1)[a]	224.6 (71.8)	.000
% Carbohydrate	50.1 (12.6)[c]	43.5 (7.7)	45 (8.8)[b]	.000
Fat (g)	52.6 (29.5)[c]	77.7 (32.6)	84.9 (28.9)[b]	.000
% Fat	33.5 (9.0)[c]	37.7 (7.4)[a]	40.8 (7.5)[b]	.000
Saturated fatty acids (g)	17.5 (11.9)[c]	23.2 (10.6)[a]	27.7 (10.6)	.000
Cholesterol (g)	242.0 (190.8)[c]	283.8 (192.1)	328.3 (185.7)[b]	.000
Dietary fiber (g)	20.1 (16.6)	20.6 (8.6)	22.2 (7.8)	.323
Vitamin A (µg)	216.5 (163.8)[c]	260.6 (142.0)	270.0 (170.0)	.001
Vitamin B1 (mg)	1.1 (0.6)	1.2 (0.5)	1.3 (0.6)	.084
Vitamin B2 (mg)	1.4 (0.7)[c]	1.6 (0.7)	1.7 (0.5)[b]	.000
Vitamin B12 (µg)	2.8 (2.1)	2.9 (1.4)	3.8 (2.6)[b]	.030
Vitamin C (mg)	244.1 (901.3)	98.3 (62.4)	112.6 (62.7)	.125
Vitamin D (µg)	3.1 (2.8)	3.2 (2.0)	3.5 (3.0)	.579
Calcium (mg)	792.7 (1304.3)	691.8 (304.2)	889.4 (966.6)	.174
Iron (mg)	10.2 (5.0)[c]	11.6 (4.4)[a]	13.3 (4.1)[b]	.000
Magnesium (mg)	253.4 (146.0)[c]	278.9 (133.4)	319.2 (328.5)	.025
Phosphorus (mg)	902.1 (422.0)[c]	1023.6 (357.5)	1108.2 (370.0)[b]	.000
Potassium (mg)	2402.6 (2162.7)	2391.2 (910.7)	2553.1 (766.2)	.560
Sodium (mg)	1729.0 (1059.2)[c]	2333.9 (1203.1)	2969.0 (1621.4)[b]	.000

[a]Post-hoc tests between Keenoa-dietitian and 3DFD with $P<.05$.

[b]Post-hoc tests between Keenoa-participant and 3DFD with $P<.05$.

[c]Post-hoc tests between Keenoa-participant and Keenoa-dietitian data with $P<.05$.

The nutrient intake between Keenoa-dietitian and 3DFD were significantly different for mean intakes of energy ($P<.001$), protein ($P<.001$), carbohydrates ($P=.001$), % fat ($P=.041$), saturated fatty acids ($P<.001$), iron ($P=.045$), and sodium ($P<.001$), with fat and cholesterol showing a statistical trend ($P=.069$ and $P=.052$, respectively).

Results of the validity analysis of the Keenoa app (Keenoa-dietitian) are summarized in Table 3 and are based on the classification method by Lombard et al [16]. The highest numbers were observed for sodium and vitamin D (n=5), and the lowest numbers were observed for vitamin B1 and phosphorous (n=1). Pearson coefficient coefficients between Keenoa-dietitian and 3DFD ranged from .38 to .51 for macronutrients and .42 to .47 for micronutrients.

Table 3. Validity analysis of the Keenoa application (Keenoa-dietitian), based on criteria levels for good (G), acceptable (A), and poor (P) outcomes.

Nutrients	Individual level			Group level			Total number of poor outcomes
	Association	Agreement		Agreement		Presence of bias	
	Pearson coefficient[a]	Cross-classification[b]	Kappa score[c]	% difference[d]	*t* test[e]	Bland-Altman[f]	
Energy (kcal)	A	P-G	P	A	P	P	4
Protein (g)	G	P-G	P	A	P	P	4
% protein	A	P-G	P	G	G	G	2
Carbohydrate (g)	A	P-G	P	A	P	P	4
% carbohydrate	A	P-G	P	G	G	G	2
Fat (g)	A	P-G	P	G	G	G	2
% fat	A	P-G	P	G	P	P	4
SFA[g] (g)	A	P-G	P	G	P	P	4
Cholesterol (g)	A	P-G	P	G	G	G	2
Dietary fiber (g)	A	P-G	A	G	G	G	1
Vitamin A (µg)	A	P-G	A	P	G	P	3
Vitamin B1 (mg)	A	P-G	A	G	G	G	1
Vitamin B2 (mg)	A	P-G	P	G	G	G	2
Vitamin B12 (µg)	P	P-G	P	P	G	G	4
Vitamin C (mg)	P	P-G	P	P	G	G	4
Vitamin D (µg)	P	P-P	P	P	G	G	5
Calcium (mg)	P	P-G	P	G	G	G	3
Iron (mg)	A	P-G	P	G	P	P	4
Magnesium (mg)	P	P-G	A	G	G	G	2
Phosphorus (mg)	A	P-G	A	G	G	G	1
Potassium (mg)	P	P-G	A	G	G	G	2
Sodium (mg)	P	P-G	P	G	P	P	5
Total number of poor outcomes	7	22 (P), 1 (G)	16	4	7	9	65
Average	N/A[h]	N/A	N/A	N/A	N/A	N/A	2.95

[a]Good, $r>.05$; acceptable, $r=.20$-$.49$; poor, $r<.20$.

[b]Good, ≥50% in the same quartile and <10% in the opposite quartile; poor, <50% in the same quartile and ≥10% in the opposite quartile.

[c]Good, ≥0.61; acceptable, 0.20-0.60; poor, <0.20.

[d]Good, 0%-10.9%; acceptable, 11.0%-20.0%; poor, >20.0%.

[e]Good, *P* value <.05; poor, *P* value ≤.05.

[f]Good, *P* value <.05; poor, *P* value ≤.05.

[g]SFA: saturated fatty acids.

[h]N/A: not applicable.

Cross-classification results revealed that the chance of misclassification was <10% for all nutrients, except for vitamin D (misclassification rate=33.8%). Weighted kappa scores ranged from 0.000 to 0.585, with an average of 0.143. Bland-Altman plots were used to compare the differences between mean intake of each nutrient between Keenoa-dietitian and 3DFD. The results showed that the majority of nutrients were in an acceptable range of agreement, with the exception of energy ($P<.001$), protein ($P<.001$), carbohydrates ($P<.001$), % fat ($P<.001$), saturated fatty acids ($P=.001$), vitamin A ($P<.001$), iron ($P=.03$), and sodium ($P=.005$).

Finally, results from the SUS showed that the mean overall score was 61.6 points (SD 19.1 points). Data were divided and analyzed by positive (questions with odd numbers) and negative (questions with even numbers) statements. The positive statement responses ranged from “neutral” to “agree,” while the negative statement responses ranged from “disagree” to “neutral.” The majority of participants believed that Keenoa

was easy to use (38/72, 52.8%) and reported they did not need the assistance of a technical person (54/72, 75.0%). There was no significant difference in acceptance between Keenoa and 3DFD (P=.28). However, 34.7% (25/72) of participants said that they would like to use Keenoa to track their diets, compared with only 9.7% (7/72) stating that they want to keep using 3DFD; 16.7% (12/72) would use both methods again, 27.8% (20/72) were not sure, and 11.1% (8/72) stated they would not use either method.

Discussion

Principal Findings

This study assessed the use of an image-based app for assessment of dietary intake in healthy adults. Similar to other studies [8,16-20], the results from Keenoa-dietitian produced similar mean nutrient profiles for more than half of nutrients when compared to the 3DFD method. However, this study included relative validation; therefore, it is impossible to conclude that one method is closer to "true dietary intake" than the other, as true dietary intake is not known.

In this study, we found significant differences between Keenoa-participant and Keenoa-dietitian data for 12 of the 22 nutrients analyzed. This suggests that the adjustments made by the dietitian were necessary to obtain the most accurate assessment of the participant's diet. Specifically, this study found that participant reports of dietary fat and protein (% difference: +31.9% and +6.9%, respectively) were lower, while the report of carbohydrates was higher (% difference: –24.7%), compared to the edited version by the dietitian. These results suggest that Keenoa is appropriate for dietitians in clinical settings; however, dietitians should review the food entries prior to generating final reports.

Despite the advantages of image-based diet-tracking apps, food identification from the user remains a challenge unless they are highly motivated to capture all food item details. An example of this lies in proper estimation of percentages of milk fat found in dairy products that is impossible to estimate from an image unless a picture of the milk carton is taken and recorded. Items that are not easily identifiable, such as milk fat from fluid milk, become problematic if diets are high in dairy-containing sauces or are included in sandwiches and other mixed-pasta dishes such as ravioli [19-22].

We speculate this to be the case in our study, as suggested by the lower reports of dietary fat and protein. Previous studies have tried to overcome this issue by inviting participants to review their image-based food diaries with a trained researcher [23,24]. However, these studies had significantly smaller sample sizes (ie, n=40 [23] and n=20 [24]) compared to our study's baseline participant pool (n=102), for which the time and resources needed to do these types of reviews were limited.

Compared with 3DFD, the energy and nutrient intakes reported via the Keenoa app (Keenoa-dietitian) were all low (average % difference, 22.5%). The highest % difference was observed for vitamin A (222%). When we excluded vitamin A from the analysis, the average adjustment decreased to 13%. These results are similar to that of other studies. In pregnant women (n=60), Savard et al [20] found an average 12.2% difference when comparing 3DFD with another web-based dietary assessment tool (24-h recall for 3 days). Similar to their study, we also compared a traditional dietary record method to a new app-based method; therefore, a relatively higher % difference is acceptable [16]. Nevertheless, in this study, the majority of the tests (4/7) scored poorly on 10 of the 22 nutrients when comparing the 3DFD and Keenoa-dietitian, including energy, protein, carbohydrates, % fat, saturated fatty acids, and iron. To our knowledge, only a few research groups have performed such an in-depth analysis as seen with the Keenoa-dietitian data, and although our findings cannot speak to reliability, the average number of poor outcomes is similar but slightly higher than those found by Savard et al [20] and Lafrenière et al [22].

It is suggested that a good Pearson coefficient should equal or surpass .5 [25]. Analysis in our study showed a weak association between 3DFD and Keenoa-dietitian since all values were <.5, except for protein. However, a relatively better association of energy and macronutrients between the two methods was observed with coefficient values closer to .5. Similarly, others have shown correlation coefficients between the 3DFD and web-based 24-h recall ranging from .03 to .76 [20]. A comparable trend was found in another study comparing a 4-day food record to two web-based 24-h recalls among 93 university-affiliated adults; the correlation coefficients varied between .06 and .76 [26]. Conversely, Lafrenière et al [22] observed a positive relative validity outcome with a mean adjusted correlation of .52 in their web-based 24-h dietary recall validation study. Notably, this research required participants to weigh their food and provide food labels or recipes, which may promote more accurate results. It also has been argued that a larger sample size could lead to a weaker correlation [19], since a good correlation (r=.46 to r=.93) was found by Wang et al [27] when studying a sample of 20 participants.

In order to assess the validity of Keenoa**,** the total number of poor outcomes based on 7 methods for each nutrient was counted. Among the 22 variables, only vitamin D and sodium had 5 poor outcomes, reflecting poor validity. Specifically, the vitamin D findings are similar to those found by others [20], which may be due to the fact that the majority of the Canadian population do not consume high volumes of vitamin D–enriched foods such as fatty fish and fortified dairy products daily [28]. In this study, vitamin D was the only nutrient with a higher rate of misclassification in the cross-classification analysis. By contrast, the higher SD of sodium intake indicates significant variability in average sodium intake, which could contribute to its considerable number of poor outcomes. This is understandable since sodium intake could vary from day to day. The average weighted kappa score was 0.143, representing slightly higher agreement and reliability between the two assessment tools at the individual level. This average is similar to that found by Landis and Koch [29] but is lower in comparison to the findings of Savard et al [20] and Lafrenière et al [22] who obtained average weighted kappa scores of 0.32 and 0.33, respectively. Overall, the outcomes at the group level (paired t tests and Bland-Altman tests) were better than those at the individual level (Pearson coefficient, cross-classification, kappa score, and % difference). These findings are supported

by those of Savard et al [20] and Lafrenière et al [22] who used the same validation tests. Therefore, the validity at the group level was stronger than at the individual level, implying that the Keenoa app is more robust at analyzing group level nutrient intakes or the general population's dietary intake. However, it should be kept in mind that bias could still exist, and the data should be interpreted carefully [16].

In this study, the average SUS score was 61.6 points, suggesting the Keenoa app "was generally well accepted; however, system users experienced usability issues" [14,30]. This is in agreement with 53.4% of participants stating that the Keenoa app was easy to use. Our results indicated that participants felt Keenoa was easy to understand and would be accepted by the general population; however, inconsistencies were found in the app, and not all participants were willing to continue using Keenoa. Indeed, compared to a written food diary, taking photos is less resource intensive, which eases the burden on participants. In line with other studies [19,31], more participants preferred using the Keenoa app over the traditional pen-and-paper method (ie, 3DFD).

Image-based dietary assessment apps, especially the Keenoa app in this study, can simplify the diet assessment process in multiple ways. Most importantly, it significantly decreases the burden on both researchers and participants by easing much of the data entry process and therefore reduces the possibility of errors [7]. In addition, because the app can be accessed remotely, it benefits people who eat out and do not take a hard copy of the food record with them [32]. Another major strength of this app is that it is linked to the Canadian Nutrient File (2015) database, which significantly reduces the data entry workload for researchers. However, some errors were still found in both Keenoa and 3DFD exported data when generating the report, which we treated as outlier data and therefore excluded those participants from this study. Although it is common to find mistakes with a new app that are often resolved with time, data should still be audited carefully in order to ensure a valid output [33]. Canada's population consists of people with different cultural backgrounds; hence, it is hard to include every food in the database. Also, estimation error is always a challenge for image-based assessment [17]; underestimation or overestimation has frequently been observed in multiple studies. Williamson et al [34] reported that estimation error was significantly decreased by employing three analysts; however, time and budget costs are a concern with this solution. Computer-aided estimation of portion sizes is an alternative direction. In the study by Fang et al [35], a machine was able to assess the portion size with a minimum range of error. Thus, estimation error may possibly be eliminated by more capable machine learning in the future.

"Technology generation" has always been a concern with the coming of new technologies [36]. Studies with younger populations were observed to have better overall outcomes and higher usability scores compared with our study [8,17,19,31,33,37]. Younger adults are more adaptable to new technologies, while elderly users are considered to have lower adoption speeds [38-40]. However, there was no significant difference found when age was included as a covariate in our research. Thus, future research is needed to expand the validity of image-based dietary assessment to all age groups, especially the elderly.

Limitations

Compared to other similar research [8,17,20,22], this study has relatively strong generalizability, since it included adults of different age groups, ethnicities, and both sexes. The diet data entry and analysis were all conducted by dietitians. However, this research has some limitations. First, our research has a relatively high dropout rate (27/102, 26.5%). One possible reason is that all participants only received a dietary report at the end of the study as compensation, which may have led to reduced willingness to cooperate. Future work may wish to assess self-efficacy or motivation to track dietary intake prior to commencing a study of a similar nature. However, the majority of participants who withdrew did not complete both methods, and the percentage of participants who failed to complete one of the two methods (ie, either 3DFD or Keenoa) was similar. This means that the dropout rate was more likely due to time or interest and was not related to the use of the Keenoa app itself. Second, the majority (58/72, 80.6%) of participants who completed the study held at least a university degree, which may have impacted their proficiency in using the mobile app [41]. In addition, one may argue that using the same 3-day dietary record for both methods would give a better estimation of validity. However, such an approach would lead to a higher workload for participants, and therefore, increase the drop-out rate and worsen compliance. Besides, nutrient intakes were reported by the participants in both methods. The use of unbiased reference measures, such as nutrition biomarkers or feeding studies, would have been the preferred method of validation against Keenoa; however, these are expensive. The possibility of systematic bias should be addressed when analyzing dietary intake based on images, and although 7 methods were used to assess the validity of Keenoa, we cannot eliminate the potential bias when interpreting the data. Finally, portion sizes of mixed items and sauces were challenging to estimate based on images, which might have contributed to the possibility of underestimating or overestimating the nutrient intake [19].

Conclusion

This study assessed the relative validity of Keenoa, an image-based mobile app, against a 3DFD in healthy adults. Our results suggest that the Keenoa app has the potential to provide accurate dietary assessment information to dietitians in a more cost-efficient and time-efficient way. Furthermore, it was well-accepted by users compared with traditional methods. However, the prediction for energy, protein, carbohydrates, % fat, saturated fatty acids, and iron intake remains questionable and should be interpreted with caution. Compared to 3DFDs, Keenoa resulted in better validity at the group level than at the individual level; thereby, it may be more effective in analyzing the dietary intake of the general population. However, a study with greater representation of older adults is needed. While participants found Keenoa easier to use compared to 3DFD, further research is needed in understanding how the app can be improved from the user perspective.

Acknowledgments

This study was funded by the R. Howard Webster Foundation through the PERFORM Centre, Concordia University (Montreal, Quebec, Canada). The authors would like to thank Dr. Stephanie Chevalier, PhD, RD for her assistance with the study design, Rainer Molla for his assistance with the study visits, and Sara Sorrini, RD for her work on the data entry and dietary analysis. The authors would also like to acknowledge Anne-Julie Tessier, RD and Anthony Grant for their assistance with the Keenoa® app.

Authors' Contributions

TC, HP, and RK conceptualized the overall study idea. TC and HP created the research questions and study design. TC and VB conducted the study and data collection. YJ analyzed the data, conducted the data analyses, and prepared the manuscript. All authors have read, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

This randomized study was not registered, as the authors specified that it did not meet the criteria deeming it necessary to formally register the study as a clinical trial based on the ClinicalTrials.gov checklist. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to their primary outcomes or effectiveness, as the lack of registration means that authors could change their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2193 KB - mhealth_v8i9e16953_app1.pdf]

References

1. Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, et al. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. Int J Epidemiol 1997;26 Suppl 1:S137-S151. [doi: 10.1093/ije/26.suppl_1.s137] [Medline: 9126542]
2. Sharp DB, Allman-Farinelli M. Feasibility and validity of mobile phones to assess dietary intake. Nutrition 2014 Nov;30(11-12):1257-1266. [doi: 10.1016/j.nut.2014.02.020] [Medline: 24976425]
3. Thompson FE, Kirkpatrick SI, Subar AF, Reedy J, Schap TE, Wilson MM, et al. The National Cancer Institute's Dietary Assessment Primer: A Resource for Diet Research. J Acad Nutr Diet 2015 Dec;115(12):1986-1995 [FREE Full text] [doi: 10.1016/j.jand.2015.08.016] [Medline: 26422452]
4. Subar A, Freedman L, Tooze J, Kirkpatrick SI, Boushey C, Neuhouser ML, et al. Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. J Nutr 2015 Dec;145(12):2639-2645 [FREE Full text] [doi: 10.3945/jn.115.219634] [Medline: 26468491]
5. Kirkpatrick SI, Vanderlee L, Raffoul A, Stapleton J, Csizmadi I, Boucher BA, et al. Self-Report Dietary Assessment Tools Used in Canadian Research: A Scoping Review. Adv Nutr 2017 Mar 15;8(2):276-289 [FREE Full text] [doi: 10.3945/an.116.014027] [Medline: 28298272]
6. Institute of Medicine. Dietary Risk Assessment in the WIC Program. In: Food-Based Assessment of Dietary Intake. Washington, DC: National Academies Press; 2002:1-184.
7. Illner A, Freisling H, Boeing H, Huybrechts I, Crispim S, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. Int J Epidemiol 2012;41(4):1187-1203. [doi: 10.1093/ije/dys105]
8. Chen Y, Wong J, Ayob A, Othman N, Poh B. Can Malaysian Young Adults Report Dietary Intake Using a Food Diary Mobile Application? A Pilot Study on Acceptability and Compliance. Nutrients 2017 Jan 13;9(1):62. [doi: 10.3390/nu9010062]
9. Turner-McGrievy GM, Dunn CG, Wilcox S, Boutté AK, Hutto B, Hoover A, et al. Defining Adherence to Mobile Dietary Self-Monitoring and Assessing Tracking Over Time: Tracking at Least Two Eating Occasions per Day Is Best Marker of Adherence within Two Different Mobile Health Randomized Weight Loss Interventions. Journal of the Academy of Nutrition and Dietetics 2019 Sep;119(9):1516-1524. [doi: 10.1016/j.jand.2019.03.012]
10. Raatz SK, Scheett AJ, Johnson LK, Jahns L. Validity of electronic diet recording nutrient estimates compared to dietitian analysis of diet records: randomized controlled trial. J Med Internet Res 2015 Jan 20;17(1):e21 [FREE Full text] [doi: 10.2196/jmir.3744] [Medline: 25604640]

11. Dietitians of Canada. Handy Serving Guide. Unlock Foods.: Dietitians of Canada URL: https://www.unlockfood.ca/EatRightOntario/media/PDFs-new-website/Portions%20Toolkit/Handy-Servings-Guide-EN-v04-July-2018.pdf [accessed 2020-04-10]
12. Health Canada. Canadian Nutrient File, 2015 version. Published February 6. 2018. URL: https://food-nutrition.canada.ca/cnf-fce/index-eng.jsp [accessed 2020-04-10]
13. Brooke J. SUS - A quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability evaluation in industry. Great Britain, UK: CRC Press; Jun 11, 1996:1-252.
14. Benyamin S, Rutter M, Smith S. The utilization of system usability scale in learning management systems: A case study of Jeddah Cummunity College. Poster presented at the 9th International Conference of Education, Research and Innovation (ICERI2016) 2016 Nov [FREE Full text] [doi: 10.21125/iceri.2016.2290]
15. Kaya A, Ozturk R, Altin GC. Usability Measurement of Mobile Applications with System Usability Scale (SUS). In: Industrial Engineering in the Big Data Era. Switzerland AG: Springer International Publishing; 2019:389-400.
16. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. Nutr J 2015 Apr 22;14(1):40 [FREE Full text] [doi: 10.1186/s12937-015-0027-y] [Medline: 25897837]
17. Howes E, Boushey C, Kerr D, Tomayko E, Cluskey M. Image-Based Dietary Assessment Ability of Dietetics Students and Interns. Nutrients 2017 Feb 07;9(2):114 [FREE Full text] [doi: 10.3390/nu9020114] [Medline: 28178196]
18. Park Y, Dodd K, Kipnis V, Thompson FE, Potischman N, Schoeller DA, et al. Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018 Jan 01;107(1):80-93 [FREE Full text] [doi: 10.1093/ajcn/nqx002] [Medline: 29381789]
19. Prinz N, Bohn B, Kern A, Püngel D, Pollatos O, Holl RW. Feasibility and relative validity of a digital photo-based dietary assessment: results from the Nutris-Phone study. Public Health Nutr 2018 Mar 6:1-8. [doi: 10.1017/s1368980018000344] [Medline: 29506585]
20. Savard C, Lemieux S, Lafrenière J, Laramée C, Robitaille J, Morisset A. Validation of a self-administered web-based 24-hour dietary recall among pregnant women. BMC Pregnancy Childbirth 2018 Apr 23;18(1):112 [FREE Full text] [doi: 10.1186/s12884-018-1741-1] [Medline: 29685127]
21. Lafrenière J, Lamarche B, Laramée C, Robitaille J, Lemieux S. Validation of a newly automated web-based 24-hour dietary recall using fully controlled feeding studies. BMC Nutr 2017 Apr 5;3(1):34 [FREE Full text] [doi: 10.1186/s40795-017-0153-3] [Medline: 32153814]
22. Lafrenière J, Laramée C, Robitaille J, Lamarche B, Lemieux S. Assessing the relative validity of a new, web-based, self-administered 24 h dietary recall in a French-Canadian population. Public Health Nutr 2018 Jul 06;21(15):2744-2752. [doi: 10.1017/s1368980018001611] [Medline: 29976261]
23. Gemming L, Rush E, Maddison R, Doherty A, Gant N, Utter J, et al. Wearable cameras can reduce dietary under-reporting: doubly labelled water validation of a camera-assisted 24 h recall. Br J Nutr 2014 Nov 28;113(2):284-291. [doi: 10.1017/s0007114514003602] [Medline: 25430667]
24. Ptomey LT, Willis EA, Goetz JR, Lee J, Sullivan DK, Donnelly JE. Digital photography improves estimates of dietary intake in adolescents with intellectual and developmental disabilities. Disabil Health J 2015 Jan;8(1):146-150. [doi: 10.1016/j.dhjo.2014.08.011] [Medline: 25281035]
25. Masson L, MCNeill G, Tomany J, Simpson J, Peace H, Wei L, et al. Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. Public Health Nutr 2007 Jan 02;6(3):313-321. [doi: 10.1079/phn2002429] [Medline: 12740081]
26. Frankenfeld CL, Poudrier JK, Waters NM, Gillevet PM, Xu Y. Dietary intake measured from a self-administered, online 24-hour recall system compared with 4-day diet records in an adult US population. J Acad Nutr Diet 2012 Oct;112(10):1642-1647. [doi: 10.1016/j.jand.2012.06.003] [Medline: 22878341]
27. Wang DH, Kogashiwa M, Ohta S, Kira S. Validity and reliability of a dietary assessment method: the application of a digital camera with a mobile phone card attachment. J Nutr Sci Vitaminol (Tokyo) 2002 Dec;48(6):498-504. [doi: 10.3177/jnsv.48.498] [Medline: 12775117]
28. Hanley D, Davison K. Vitamin D insufficiency in North America. J Nutr 2005 Feb;135(2):332-337. [doi: 10.1093/jn/135.2.332] [Medline: 15671237]
29. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310] [Medline: 843571]
30. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
31. Simpson A, Gemming L, Baker D, Braakhuis A. Do Image-Assisted Mobile Applications Improve Dietary Habits, Knowledge, and Behaviours in Elite Athletes? A Pilot Study. Sports (Basel) 2017 Aug 11;5(3):60 [FREE Full text] [doi: 10.3390/sports5030060] [Medline: 29910420]
32. Forster H, Fallaize R, Gallagher C, O'Donovan CB, Woolhead C, Walsh MC, et al. Online dietary intake estimation: the Food4Me food frequency questionnaire. J Med Internet Res 2014 Jun;16(6):e150 [FREE Full text] [doi: 10.2196/jmir.3105] [Medline: 24911957]

33. Rangan AM, Tieleman L, Louie JCY, Tang LM, Hebden L, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): relative validity of a mobile phone application to measure intake of food groups. Br J Nutr 2016 Jun;115(12):2219-2226. [doi: 10.1017/S0007114516001525] [Medline: 27121045]
34. Williamson DA, Allen H, Martin PD, Alfonso AJ, Gerald B, Hunt A. Comparison of digital photography to weighed and visual estimation of portion sizes. Journal of the American Dietetic Association 2003 Sep;103(9):1139-1145. [doi: 10.1016/s0002-8223(03)00974-x]
35. Fang S, Liu C, Zhu F, Delp E, Boushey C. Single-View Food Portion Estimation Based on Geometric Models. 2015 Nov Presented at: 2015 IEEE International Symposium on Multimedia; 2015; Miami, Florida p. 385-390. [doi: 10.1109/ism.2015.67]
36. Paul G, Stegbauer C. Is the digital divide between young and elderly people increasing? First Monday 2005 Oct 03:3 [FREE Full text] [doi: 10.5210/fm.v10i10.1286]
37. Shoneye C, Dhaliwal S, Pollard C, Boushey C, Delp E, Harray A, et al. Image-Based Dietary Assessment and Tailored Feedback Using Mobile Technology: Mediating Behavior Change in Young Adults. Nutrients 2019 Feb 19;11(2). [doi: 10.3390/nu11020435] [Medline: 30791502]
38. Casperson SL, Sieling J, Moon J, Johnson L, Roemmich JN, Whigham L. A mobile phone food record app to digitally capture dietary intake for adolescents in a free-living environment: usability study. JMIR Mhealth Uhealth 2015 Mar 13;3(1):e30. [doi: 10.2196/mhealth.3324] [Medline: 25775506]
39. Hongu N, Pope BT, Bilgiç P, Orr BJ, Suzuki A, Kim AS, et al. Usability of a smartphone food picture app for assisting 24-hour dietary recall: a pilot study. Nutr Res Pract 2015 Apr;9(2):207-212 [FREE Full text] [doi: 10.4162/nrp.2015.9.2.207] [Medline: 25861429]
40. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012 Apr 13;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
41. Robinson J, Neustadtl A, Kestnbaum M. The Online "Diversity Divide": Public Opinion Differences Among Internet Users and Nonusers. IT & Society 2002;1(1):284-302 [FREE Full text]

Abbreviations

3DFD: 3-day food diary
A: average
FFQ: food frequency questionnaire
G: good
P: poor
SFA: saturated fatty acid
SUS: System Usability Scale

Edited by G Eysenbach; submitted 26.11.19; peer-reviewed by L Catherine, R Abd Jalil; comments to author 15.03.20; revised version received 06.05.20; accepted 13.06.20; published 09.09.20.

Please cite as:
Ji Y, Plourde H, Bouzo V, Kilgour RD, Cohen TR
Validity and Usability of a Smartphone Image-Based Dietary Assessment App Compared to 3-Day Food Diaries in Assessing Dietary Intake Among Canadian Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(9):e16953
URL: https://mhealth.jmir.org/2020/9/e16953
doi:10.2196/16953
PMID:32902389

Original Paper

Design and Development of a Digital Weight Management Intervention (ToDAy): Qualitative Study

Charlene L Shoneye[1], BSc, Grad Dip, PGD, MSc; Barbara Mullan[2], BA (Hons), MA, PhD; Andrea Begley[1], BSc, Grad Dip, MPH, DrPH; Christina M Pollard[1,3], BSc, Grad Dip, MPH, PhD; Jonine Jancey[4], BSc, PhD; Deborah A Kerr[1], BSc, Grad Dip, MSc, PhD

[1]School of Public Health, Curtin University, Perth, Western Australia, Australia

[2]Health Psychology & Behavioural Medicine Research Group, School of Psychology, Curtin University, Perth, Western Australia, Australia

[3]East Metropolitan Health Service, Perth, Western Australia, Australia

[4]Collaboration for Evidence Research & Impact in Public Health, School of Public Health, Curtin University, Perth, Western Australia, Australia

Corresponding Author:
Deborah A Kerr, BSc, Grad Dip, MSc, PhD
School of Public Health
Curtin University
GPO Box U1987
Perth, Western Australia
Australia
Phone: 61 61892664122
Email: D.Kerr@curtin.edu.au

Abstract

Background: The Tailored Diet and Activity (ToDAy) study aims to build on the campaign by adding a digital intervention with the potential to provide wide-reaching, cost-effective weight management support.

Objective: The ToDAy study aims to build a tailored intervention using mobile technology to improve diet and physical activity behaviours in adults with overweight and obesity. The main objectives were to identify behavior change techniques for diet and physical activity (PA) change for weight loss and explore preferences for digital intervention features that would be effective in changing diet and PA behaviors.

Methods: This qualitative study uses the principles of a person-based approach to intervention development; the behavioral intervention technology framework; and the capability, opportunity, motivation, and behavior (COM-B) framework. Focus groups and telephone interviews were conducted with 56 adults in Western Australia. Open-ended questions and example intervention features were used to explore the usability and acceptability of the self-monitoring tools, knowledge about effective weight-loss strategies, and acceptability of tailored feedback. Findings from the focus groups and interviews were analyzed using thematic analysis.

Results: Qualitative findings revealed an awareness of key public health messages but a lack of confidence in how to perform these behaviors to help manage their weight. A total of 4 major themes were identified and mapped to the domains of the COM-B framework: (1) misinformation, (2) environmental support, (3) social norms, and (4) confidence.

Conclusions: This study explores users' capability, opportunity, and motivation to perform the target behaviors for weight loss. The findings suggested that a digital weight management intervention using a mobile food record and activity trackers to inform tailored feedback may be acceptable and feasible. Participants expressed a preference for simple expert advice, digital self-monitoring tools, and visual feedback.

International Registered Report Identifier (IRRID): RR2-10.2196/12782

(JMIR Mhealth Uhealth 2020;8(9):e17919) doi:10.2196/17919

KEYWORDS

obesity; diet; physical activity; sedentary behavior; digital behavioral interventions; health behavior; wearable activity monitor; health; mobile food record; clinical trial; focus group; qualitative research; mobile phone

Introduction

Background

Excess weight has overtaken smoking as the leading cause of noncommunicable disease in Australia, with 7 out of 10 males and almost 6 out of 10 females living with overweight or obesity [1]. The causes of this are multifaceted [2] but, at a personal level, poor diet and inactivity are major contributors. In Australia, excessive intake of alcohol, sugar-sweetened beverages (SSBs), and *discretionary foods* (foods considered to be of little nutritional value; often high in saturated fats, added sugar, and salt; and alcohol or *junk* foods) are observed across all age groups [3]. Mass media campaigns, targeting healthy weight, positively influence knowledge and awareness with modest impacts on behavior [4-7]. LiveLighter is a Western Australian public health education and social marketing campaign that aims to encourage people to eat well, be physically active, and maintain a healthy weight. The campaign engages with the community through paid and unpaid social media, web-based and printed resources, and retailers. Campaign messages include graphic images of *toxic* fat, followed by messages with single actions to reduce the risk of weight gain, for example, by avoiding SSBs or junk food [8]. The advertisements also direct people to a campaign website where there is an option to enroll on the web to access a meal planner, recipes, and weight-monitoring tools and to receive update emails. The Tailored Diet and Activity (ToDAy) study aims to build a digital intervention that provides individualized tailored feedback on dietary and activity behaviors.

In Australia, the evaluation of the LiveLighter mass media campaign targeting sugary drinks indicates a high campaign recall and modest reductions in SSBs [4,9-11].

However, >60% of adults in Australia do not usually consume SSBs and may disregard the campaign, even if they stand to benefit from some of the other elements. A growing body of evidence supports the notion that information is best tailored specifically to the unique characteristics and behaviors of an individual [12-17], with significant effects reported for nutrition and PA.

Australian clinical guidelines for weight management recommend a multidisciplinary team of health professionals using specific behavior change techniques (BCTs) applied for a minimum of 12 months [18]. High attrition rates, low availability of trained health professionals, and logistical issues are among the commonly reported barriers to putting this in practice [19]. Digital interventions that combine clinical and tailored content with the reach of mass media could be a cost-effective solution [20].

Developing Digital Interventions

The process of developing a digital intervention requires integrating the behavior change theory with intervention features that are engaging and acceptable to the target group. Evidence from other digital behaviour change interventions suggest an iterative and multidisciplinary approach that includes a qualitative investigation with the end user before implementation [14,21,22]. Specifically, the person-centered approach recommends using qualitative research to explore and test intervention features with the target group [23]. This allows researchers to adapt the intervention features based on the preferences and needs of the user. A BCT is the *active ingredient* or intervention component that changes the desired behavior [24]. For example, BCT 2.3 is self-monitoring of behavior, the most commonly used BCT in effective diet and PA interventions [25]. Evidence suggests that self-monitoring works through behavioral regulation; for example, dietary self-monitoring increases the awareness of food choice, portion sizes, and improving diet quality [25,26,27]. Exploring digital interventions for weight management that combine the reach of mass media campaigns with tailored and clinical support could be a cost-effective and practical approach.

Focus groups and interviews are commonly used to generate discussion and explore participants' experiences as well as their needs, knowledge, and preferences [28,29]. However, weight stigma inhibits participation and the willingness to share personal beliefs [29,30]. One strategy to address this is to show participants hypothetical scenarios and ask them to provide advice for weight loss [28,29]. Another strategy to address this is to use hypothetical scenarios where information about a person's PA levels or images of their meals is shown and participants are asked to provide advice on weight loss. To date, this unique approach has not been undertaken in this population.

Although men are more likely to be living with excess weight and experience health-related illnesses, they are less likely to participate in lifestyle interventions, making up only about 20% of participants [31]. Studies aiming to reduce this gender imbalance have reported that self-monitoring technology, including mobile apps and wearable devices, are great incentives to engage men in weight management [32,33]. To date, few weight management studies have included the views and experiences of men aged >25 years [34]. This study aims to address this shortcoming by purposely sampling an equal number of male and female participants.

This study aims to describe the qualitative study and iterative process used to develop ToDAy, a digital, tailored, weight management intervention. A full description of the aims of the 12-month intervention and the protocol has been published elsewhere [35]. The objectives of this study were to (1) identify BCTs for dietary and PA changes concerning weight loss and (2) explore preferences for digital intervention features that would be effective in changing diet and PA behaviors.

Methods

Study Design

The methodological approach used in this research was a general inductive, qualitative approach [36]. The Consolidated Criteria for Reporting Qualitative Research for interviews and focus groups were used to ensure rigor in the presentation of the findings [37]. ToDAy will incorporate learnings from an earlier trial where a mobile food recording app (mFR) was successfully used to assess dietary intake and provide tailored feedback on fruit, vegetable, and junk food intake in young adults [38,39].

Open-ended questions and example intervention features were used to explore the usability and acceptability of the self-monitoring tools, knowledge about effective weight loss strategies, and acceptability of the tailored feedback.

We used an iterative intervention development process applying the behavioral intervention technology (BIT) theory [40,41] and the person-based approach [23]. An overview of the 3 stages of intervention development is provided below.

Stage 1

To derive the measurable and clinically significant behavioral changes that could be expected from the intervention, we reviewed the recent literature and evidence-based guidelines from the Australian scientific authoritative bodies. These included the National Health and Medical Research Council's clinical guidelines for weight management [18], the Australian Dietary Guidelines [42], and the Australian government's PA guidelines [43]. We determined the clinically significant target behaviors for the intervention as follows:

Dietary:

1. Daily dietary energy reduction of 2000 kJ.
2. Avoiding or limiting energy-dense nutrient-poor (EDNP) foods, SSBs, and alcohol.
3. Eating less at meals or additional snacks (except fruits and vegetables).
4. Eating less often [18,42].

PA:

1. Daily step count ≥10,000.
2. ≥30 active minutes (spent in moderate-to-vigorous PA).
3. ≥250 steps per hour [43].

Weight loss:

1. 5% reduction in body weight [18].

Target behaviors: The ToDAy study investigated whether a digital, tailored intervention can improve diet and PA behaviors in adults with overweight or obesity. A total of 16 health professionals with expertise in dietetics, PA, health promotion, and community engagement were consulted in a series of 5 workshops and meetings to explore the target behaviors needed to achieve the clinical aims and where, when, why, and who they occur with [45].

Stage 2

The research team developed a user-friendly script for the focus groups and interviews with a male and female consumer representative [46]. Focus groups and interviews were conducted with volunteers to explore the acceptability of the selected BCTs and their preferences for digital intervention features. The findings of these focus groups are presented in this paper.

Stage 3

Target behaviors were mapped to possible intervention features by the research team with reference to previous research [47,48] and following guidelines for developing complex behavior change interventions [49,50]. Focus groups and interviews were followed by a review of intervention features by the research team. This was repeated in a cyclical manner to allow continued user involvement in the design and development of the final intervention.

Approval for the study was granted by the Curtin Human Research Ethics Committee (HR E2016-0271). All participants agreed to an audio recording of their focus group or interview and provided informed consent. All data were collected between October and November 2016 in Western Australia (spring).

Theoretical Frameworks

Several guidelines exist for the development and assessment of evidence-based apps and web-based interventions [50-52]. As this intervention uses a combination of digital tools, that is, an mFR, a wearable PA tracker, text messages, and emails, a combination of theoretical approaches and guidelines was drawn upon. The BIT framework was used to identify the technology and procedures for delivering clinical aims and BCTs (objective 2) [40]. The models help to identify clinical aims and link these with suitable intervention features for testing with the user (Table 1).

The capability, opportunity, motivation, and behavior (COM-B) model was then used to guide the selection of intervention features and strategies such as self-monitoring, goal setting, motivation enhancement, and feedback on performance (objective 1) [45]. The COM-B model aims to specify behavioral targets and support psychological theories when developing interventions [45]. The COM-B model states that 3 factors are needed to change behavior: capability (C), opportunity (O), and motivation (M). According to this model, performing a behavior (B) first requires individuals to be capable (C) or have physical and mental abilities (eg, nutrition knowledge, cooking skills). Following this is opportunity, which includes both practical and social aspects (eg, access to healthy food that is culturally acceptable and within social norms). Finally, motivation includes automatic drivers like habits as well as beliefs, plans, and impulses. Table 1 illustrates the steps in the development process—why, how, what, and where? The *what* includes the BCT and associated taxonomy number to identify each BCT from the Behavior Change Technique Taxonomy v1 (BCTTv1) [53].

Table 1. Relationship among clinical aims, behavior change techniques, and intervention features (technology).

Why? Clinical aim or population health focus	How? Action	What: behavior change techniques[a] [53]	Where: potential intervention features tested in qualitative research
Reduce BMI by 5% [18]	Reduce energy intake by 2000 kJ per day and increase PA[b] (10,000 steps)	Provide information on the consequences (5.1), goal setting (behavior and outcome; 1.1, 1.3), and review of behavior goals (1.5)	Tailored feedback, weight tracker, PA tracker
Reduce EDNP[c] foods [42]	Increase awareness of EDNP intake	Goal setting (behavior and outcome; 1.1, 1.3), review of behavior goals (1.5), provide feedback on behavior (2.2), and social comparison (6.2)	Mobile food record; tailored feedback and tailored education; app alerts, eg, Have you had any snacks today?
Reduce SSBs[d] [42]	Increase awareness of energy in SSBs and intake	Self-monitoring of behavior (2.3), goal setting (1.1), barrier identification, provide feedback on behavior (2.2), and social comparison (6.2)	Mobile food record tailored feedback and tailored education
Increase fruit and vegetable consumption [42]	Increase awareness of current intake	Self-monitoring of behavior (2.3), discrepancy between current behavior and recommendations (1.6), action planning (1.4), problem solving (1.2), and instruction on how to perform behavior (4.1)	Mobile food record tailored feedback and tailored education
Reduce alcohol intake [54]	Increase awareness of current intake	Information on health consequences (5.1), motivational interviewing, and self-monitoring of behavior (2.3)	App alerts, eg, How confident are you about having an alcohol-free dinner tomorrow night?

[a]Behavior change technique and associated taxonomy from the Behavior Change Technique Taxonomy v1 (BCTTv1) [53].

[b]PA: physical activity.

[c]EDNP: energy-dense nutrient-poor.

[d]SSBs: sugar-sweetened beverages.

Recruitment

Recruitment was specific and purposeful [55], aiming for a similar number of males and females and including people with overweight or obesity who had some experience of the LiveLighter campaign [8]. A single recruitment email was sent to 20,000 adults who had registered with the LiveLighter website in October 2016. The email was sent to the entire mailing list of LiveLighter members, inviting them to take part in the study by clicking on a study web link where participants completed web-based consent and screening. The website was closed after 2 days as 245 respondents had completed the screening questionnaire. The respondents who met the criteria were sent further details on how to participate in a focus group or interview. There were 145 eligible participants, who were >18 years and had a BMI over 25 kg/m^2. The time, date, and location of the focus groups were sent to the eligible participants to allow them to choose the most convenient time. As 85% of the sample were women, additional recruitment strategies were employed to encourage male participants, such as offering a one-to-one telephone call and men-only focus groups. When these additional approaches were not successful, a workplace with a high proportion of males was contacted and an onsite focus group was arranged, with 14 men in attendance.

Script Development

The topics covered in the focus groups and interviews were informed by the literature as important features for weight loss interventions and included self-monitoring of diet and PA behavior, feedback on performance, reducing intake of discretionary food drink and alcohol, reducing sedentary time, and increasing steps per day. The script was pilot tested with researchers at Curtin University and 2 consumer representatives where feedback on clarity was provided. Focus groups were conducted in community settings, community centers, place of work, and Cancer Council WA meeting rooms. For each session, the script was accompanied by a visual presentation of example images and draft intervention features. Multimedia Appendix 1 contains the basic script used for focus groups and interviews.

A semistructured focus group and interview guide with open-ended questions were developed, which allowed an iterative, person-centered data collection process [23,56]. As a result, a variation of the script was used in each session. For example, “the last group suggested the dietary feedback include their food images so they can see where the junk food came from. What do you think of this idea? Have a look at this example. Is there anything you would change?”

User Preferences

Preferences for digital intervention features explored willingness to use the digital self-monitoring tools as well as the frequency and duration of self-monitoring. Preferences for digital feedback explored the format, frequency, length, and content of the tailored diet and PA feedback. With regard to digital content to address the target behaviors, participants were asked to suggest helpful advice and potential barriers to changes for a particular behavior, for example, “What feedback could we send to help this person lose weight?” and “What sort of things do you think might get in the way?”

Self-Monitoring Tools

The usability and acceptability of the self-monitoring tools were explored. First, participants were given an opportunity to use the mFR [57-59]. This image-based dietary assessment tool uses the integrated camera in a smartphone to capture images of food and beverages. The images are automatically uploaded

to a server for dietary analysis by the research dietitian and used to inform the tailored dietary feedback. The usability and acceptability of the mFR to monitor dietary intake and a wearable device to monitor PA were explored.

Behavior Change Beliefs Regarding Weight Loss

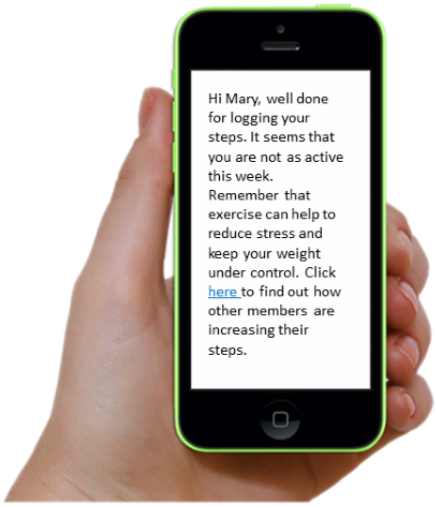
Participants were given examples of a scenario and asked to provide advice to a hypothetical person to help them lose weight. For example, "people in this study will use an app on their phone to take pictures of their food and drink. Imagine we received this picture from a man wanting to lose weight, what advice should we give him?" (Figure 1). Participants, who were interviewed on the phone, were emailed this information and the images in advance.

Figure 1. Example of image shown to participants where they were asked what advice they would give this person to help them lose weight.

Acceptability of the Feedback Messages

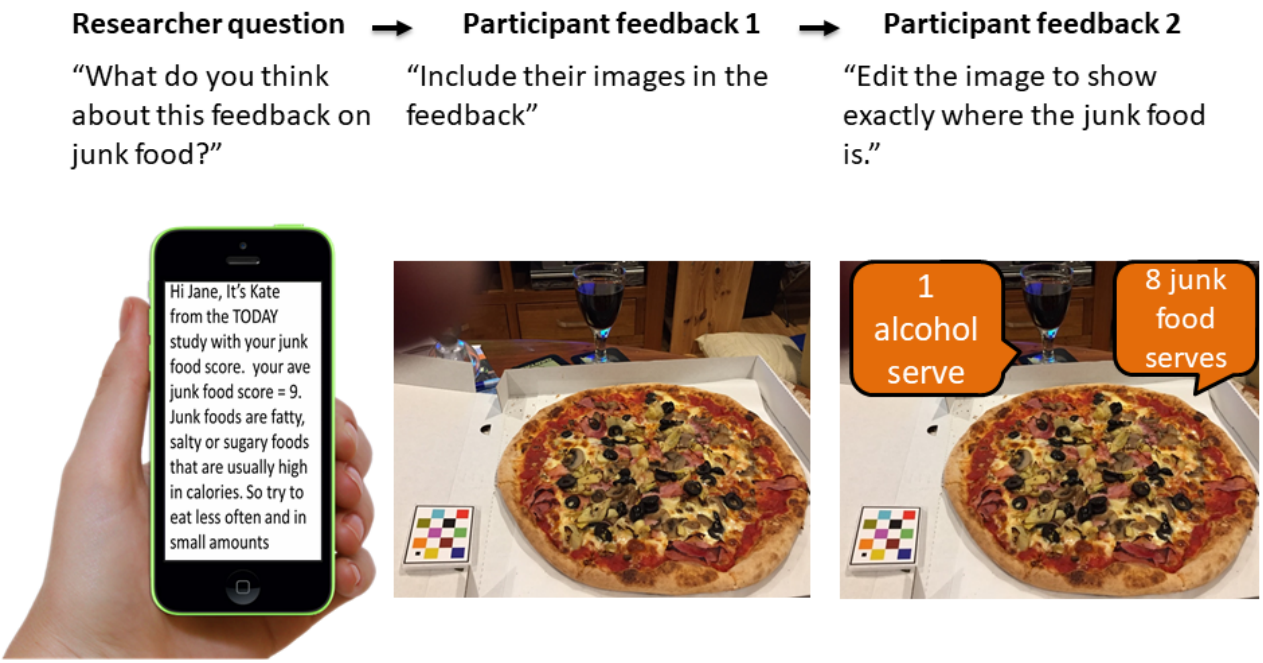
The acceptability of feedback messages, including the length, content, and tone, was explored. Example feedback on diet and PA behavior were shown to participants with questions to explore their understanding and acceptability of the feedback. For example, Figure 2 shows an example of PA feedback where participants were asked to "imagine we sent a person this feedback. How do think they might feel about receiving this feedback? Is there any other information you would add?"

Figure 2. Example of physical activity feedback shown to participants for their comment.



After each focus group, new ideas were discussed with the research team (qualitative researcher, dietitian, exercise physiologist, and health psychologist) and potential digital intervention features were developed. The script was adapted after each session using an iterative process to incorporate participant ideas and feedback, which were then explored in the subsequent sessions. Figure 3 provides an example of how intervention features evolved using feedback from participants.

Figure 3. Example images showing how new ideas on how to provide dietary feedback were incorporated using an iterative process based on feedback from the previous focus group and interviews.



Data Collection

All interviews were conducted over telephone by author 1, a dietitian (CS; female) who has qualitative research experience [60], with guidance from AB, an established qualitative researcher [61,62]. Participants in the interview completed a consent form and a demographics questionnaire on the web. The focus group participants completed a consent form and a demographic questionnaire on arrival. JH facilitated the male-only group, with CS as a cofacilitator. CS facilitated all other groups, with 1 assistant moderator. Before the interviews and focus groups, participants had no relationship with the researchers and knew the study was about helping to develop a digital weight management intervention.

All focus groups and interviews started with an overview of the proposed ToDAy intervention, where participants monitor their PA with a wearable tracker and record their food and beverage intake with the mFR app. This information was used to provide feedback to help them lose weight. The first activity was a chance to employ the mFR used to capture images of food and beverage intake [57,58]. The facilitator demonstrated how to use the app to take pictures of plastic food models. Participants were then given an opportunity to use the mFR on a mobile device. This exercise served as an icebreaker as well as capturing the questions and comments of participants using the app. This was followed by open-ended questions to start the discussion. Focus groups and interviews lasted between 34 and 78 min and were conducted until reaching a saturation of ideas. At the end of each focus group, the facilitator summarized the main ideas or themes that participants had raised in that group and gained agreement from participants.

Data Analysis

All audio recordings were professionally transcribed verbatim and reviewed for accuracy by the first author and managed in NVivo. As this study used scripts that evolved between groups, the analysis used a thematic analysis to analyze and code data through the lens of the COM-B model [49]. Qualitative data were analyzed in 3 stages. Following the process of thematic analysis, first was familiarization through reading each transcript, highlighting key points, and discussing the findings with the cofacilitators [63]. The first author led the analysis and developed themes aligning with capability, opportunity, and motivation. The cofacilitators for the focus groups then independently reviewed the scripts. Finally, any discrepancies were reviewed and discussed by the first author. The quotes were then aligned to the final themes [64].

Results

Participant Characteristics

Table 2 provides an overview of the characteristics of the participants. A total of 56 adults (32 female and 24 male) from Western Australia participated in 6 telephone interviews and 6 focus groups (average of 5 per group). Of these, over one-third had a BMI>30. Half were employed full time (52%), and most were aged between 25 and 40 years (61%). All participants owned a smartphone (iPhone or Android) and had some experience using apps.

Feedback from focus groups and interviews provided important insights into the acceptability and comprehension of tailored feedback messages. For example, feedback suggesting healthier alternatives to junk food was rejected as these options may not be available and could not be tailored to the individual's preferences.

Table 2. Characteristics of interview participants (N=56).

Characteristics	Values, n (%)
Age (years)	
25-40	34 (61)
41-65	17 (30)
>65	5 (9)
Gender	
Male	24 (43)
Female	32 (57)
BMI (kg/m^2)	
<25	10 (18)
25-30	23 (41)
>30	23 (41)
Ethnicity	
Australian	47 (84)
Indigenous Australian	5 (10)
Asian	4 (7)
Employment status	
Employed full time or part time	39 (70)
Unemployed	8 (14)
Retired	9 (16)
Household income Aus $ (US $)	
<$50,000 (<$35,695)	21 (38)
$50,000-$150,000 ($35,695-$107,085)	22 (40)
>$150,000 (>$107,085)	13 (22)

Qualitative Analysis

Emerging qualitative themes and subthemes were mapped to the COM-B domains. This helped to identify that participants needed support in all 3 areas of capability, opportunity, and motivation.

Capability: Misinformation

Capability refers to knowledge and skills related to behavior [45]. Participants' knowledge of weight loss behaviors was examined by asking them to provide dietary advice to a hypothetical client based on a picture of their food and drink. Responses were themed as misinformation when they provided inaccurate information or nutrition advice. The majority of the discussion focused on giving misinformation as dietary advice. This revealed their knowledge and beliefs about which behaviors are best for weight management. For instance, 2 main examples were discussed. First, potatoes were considered fattening and second, excess alcohol was not a major contributor to weight gain. There was a focus on individual food being responsible for weight gain rather than a holistic view of the total diet.

Carbohydrates Cause Weight Gain

Participants were shown an extra-large roast dinner with large meat portions, 3 small potatoes, 3 serves of vegetables, and 6 bottles of beers and asked, "What advice would you give this person to help them lose weight?" All groups commented on the potatoes, only later mentioning the large serve of meat and 6 bottles of beer:

> *It's not a healthy meal if it's got a potato.* [Female, FG2]
>
> *Cut down on the carbs, only 2 potatoes.* [Male FG3]

Alcohol Consumption Does Not Cause Weight Gain

The discussion did eventually focus on reducing the 6 bottles of beer, but there were a number of misinformed views about alcohol situations where 6 beers would be *OK*. For example, if it was a low-carb beer, consumed with fresh lemon, watching football, consumed over the course of an afternoon, the participant has eaten well during the preceding week or has a physical job that would work off excess energy. Some participants were dismissive of the kilojoule content of alcoholic drinks but aware that alcohol may lead to choosing discretionary foods:

> *So the kebab or the burger is a lot more appealing after a few drinks than it may be if say, while sober.* [Male, PI 1]

It's not the beer itself that's the problem, it's the food you have with it. [Female, FG2]

Opportunity: Environmental Support and Social Norms

The *opportunity* component of COM-B relates to physical opportunities, such as the environment and availability as well as social opportunities, including social influences. Both emerged as important themes in the data.

Lack of Environmental Support

Environmental factors were discussed as the main barriers to avoiding junk food. One group, in particular, expressed frustration at the density of fast food outlets and the promotion of very low-cost meals that would appeal to children and those on a low income:

When I first came in 1982 there was hardly any fast food places. Now you've got 8 or 9 different ones all down the road from each other. Kids have got 2 dollars where are they going to go? For a $2 burger. [Male, FG 5]

Participants expressed that the social marketing campaigns and interventions for individuals were futile without addressing the food environment:

…that goes back to the government again because the government has put no restrictions on how much fast food can be in local vicinities. If you look at where they are, it's not in the high-class areas it's in the low socio-economic suburbs. [Male, FG 5]

Unhealthy Food Is Everywhere

Many participants thought that most food options purchased outside the home were not healthy and the serving sizes were too large. Some mentioned added fat, sugar, and salt. Others noted that savory meals often come with chips and sweets are served with cream or ice cream:

When you buy food out... It all tends to be fattening. [Female, FG 2]

Most groups brought up opportunistic eating as a key facilitator of eating discretionary food. Events such as at sporting occasions, bake sales, and *sausage sizzles* were given as common examples:

You can't buy a healthier option at a sausage sizzle, they only have sausage, white bread, and sauce. [Male, FG 4]

Social Norms

There has been some discussion about the difficulty of social opportunities. Most participants agreed that the expectations of consuming junk food at social occasions were problematic. When eating out, both men and women agreed that it is not socially acceptable for men to ask for healthy options or modification to their order. Regarding swapping chips for vegetables, a man said:

It's not seen to be manly to be seen like, eating vegetables. [Male, FG 3]

If it comes with chips, I'll eat chips. [Male, FG 4]

When discussing ways to reduce alcohol consumption, participants suggested practical methods, such as alternate alcohol with sugar-free soft drinks or choosing a low-strength beer. However, there was an overwhelming consensus that these suggestions were unrealistic and not socially acceptable for men:

…might be difficult if you're down at the club or in the pub with the guys and then you really get the Mickey taken out of you [Male, FG4]

You could suggest having water for every other drink, but who is going to listen to that? [Male, FG 5]

Motivation

The main theme that emerged from the motivation domain of COM-B was confidence.

Confidence

Participants were uncertain about their knowledge or beliefs about food, PA, or weight management. Some of this was linked to the perception that the guidelines from professionals were *always changing*. There was a sense of complacency about the need to or the importance of achieving health recommendations:

What would have been recognized as a healthy meal years ago, these days it's not a healthy meal because it's got potato; carbs. [Male, PI 3]

Already Doing Enough

In relation to PA, most adults felt they were already active with daily activities, including looking after children, gardening, housework, and those in nonsedentary employment (eg, nurse, carpenter, and plumber). Motivation to engage in PA was limited by the belief that their lives were already active and busy:

... A mother of 3, busy all day, being told to go for a 45-minute walk. They can't otherwise they would have done it. [Female FG 2]

Concerning eating, most agreed that there was a place for junk foods, namely takeaways, confectionery, and desserts. Making healthy food choices was said to be important, especially for those with health problems like diabetes and high blood pressure. There was a variety of beliefs about how often people should make healthy food choices. Some used broad terms, such as sometimes or not too often. Others believed that healthy eating and being active were part of the working week and not applicable on the weekend:

I mean you can eat healthy but every now and again you can always have take away. But not every week or every day. If you know what I mean? [Female FG 1]

During the week it's structured. You've got, you know, time to get to work and your lunch break at work or whatever and then, you know, you come home and it's dinner time and then that's that. Then your weekends are your time to just flop. It's like a treat. Weekends are your time to relax and enjoy life I guess. [Female, PI 2]

Reality of Change

When the group was shown potential feedback to address alcohol intake, there was general agreement that changing would be difficult or even impossible. Specifically, several people thought that any feedback to reduce alcohol intake would be futile:

> *You could tell him to stop at four beers, but by then he won't know what he's doing.* [Male, FG 5]

Ambivalence

There was a conflict in the expectations of the interventions. Some felt that feedback should tell them the negative consequences of their poor health choices:

> *In three years at that rate, your liver is going to look like this.* [Male, FG 5]

However, feeling judged or reprimanded was cited as a barrier to keeping people engaged and honest about self-reporting. Participants wanted a specific example of food they could swap rather than a general *avoid this* or *chose a healthier option*. At the same time, they said specific advice would be unrealistic:

> *All well and good to suggest a healthy option but you can't get your chicken parmigiana grilled with a baked potato at my local, it comes deep-fried with chips.* [Female, FG 3]

These comments reflect the ambivalence about eating junk food outside the home; something they want help to avoid because it is expensive, is unhealthy, and leads to overeating but something they do because it is convenient and enjoyable.

Functions and Features of the Intervention

Results from the focus groups and interviews informed the selection of user preferences, intervention functions, and acceptability of the intervention. Table 3 shows the results on the acceptability of digital tools proposed for self-monitoring diet and PA. Participants found the mFR app to be intuitive and convenient in comparison with other digital tools or paper-based methods. All participants shared experiences of using a pedometer and saw the option for a wearable PA tracker as an incentive to join the study. Feedback on the clinical aims highlighted gaps in their understanding of the guidelines for diet and PA. Participants agreed that personalized feedback would promote health-enhancing habits by enhancing confidence and motivation.

The focus groups and interview data were reviewed for 3 intervention functions expected to mediate a behavior change, for example, education, modeling behavior, and persuasive communication.

1. *Education* (to increase knowledge on how to identify EDNP food and on PA guidelines)
2. *Modeling behavior* (annotating food and beverage images by feedback on intake)
3. *Persuasive communication* (images and motivation enhancement using positive reinforcement in tailored communications)

Table 3. Understanding user perspectives and experiences of participants on the clinical aims using the mobile food record app and their experiences or views about using a physical activity monitor.

Clinical aims and examples of questions or activity	Participant quotes
Self-monitor diet	
Practice using the mobile food record app	"It's really intuitive and easy; better than the apps where you need to find the food"
How easy/difficult would it be to use to capture all your food and drink for 4 days?	"I wouldn't use it if I was at the club with the guys. I wouldn't use it at work (nurse)"
Self-monitor physical activity	
Have you ever used an activity tracker? Prompt for wearable device, pedometer, mobile app	"I used to use a watch that tracked steps and heart rate, it was good at first then all the alerts got annoying"
Any advice or support that helped?	"Yeah it's good to see that you've done like ten thousand steps in a day"
Increase fruit and vegetable consumption	
How much fruit and vegetables are recommended each day?	"2 fruit and 5 veg but I'm not sure if it has to be 5 different types"
What advice would you give this person to help them lose weight? (Shown example meal)	"your vegetables are supposed to be half your plate so many people don't know that"
What type of feedback could we send to help this person to get them to eat more vegetables?	"If someone sends you a picture, send it back saying 'that's at least one serve of your five today'"
What sort of things might get in the way?	"No one eats that much veg, it's impossible"
Reduce EDNP[a]	
What advice would you give this person, to help them lose weight?	"because you've eaten this you have to run ten kilometers to work it off, sort of thing"
What sort of things might get in the way?	"…you go to Bunnings (national hardware chain with fundraising barbecues) you're going to get your sausage, there's no other options"
Reduce intake of alcohol	
What advice would you give this person, to help them lose weight?	"Would be much better if he switched to a lighter beer, like only 4%"
What sort of things might get in the way?	"If you're out, say watching sport in the afternoon, everyone else is drinking it would be hard to have water"
Increase active minutes/decrease sedentary behavior	
How much physical activity is recommended each day?	"The adverts tell you half an hour a day and ten thousand steps. You can't do ten thousand steps in half an hour, it doesn't make sense"
Any ideas of how we could support them to do this?	"Show it like your bank account where you can track it and see where it's going"
What might make it easier/difficult?	"It needs to be friendly and informative and helpful"

[a]EDNP: energy-dense nutrient-poor.

Table 4 summarizes the stages of applying COM-B to the focus groups and interview data to identify nutrition, PA, and weight management intervention functions and features.

New ideas from participants were explored for feasibility and consistency with evidence-based diet and PA guidelines [42,43]. Rejected ideas included individual assessment of vitamin and mineral status and details on how much weight could be lost with a specific number of steps, as this was not consistent with evidence-based advice. Accepted ideas were the inclusion of participant food and beverage images from the mFR into the tailored dietary feedback. For PA, accepted ideas included using graphs for self-comparison of PA levels throughout the study, positive re-enforcement to acknowledge improvements in activity, and regular goal setting. There was a strong preference for visual feedback for both diet and PA. This was not feasible with text messages; therefore, feedback was primarily provided by email. Digital intervention features may continue to develop during the intervention and include aspects that were not originally mapped in this development phase [40].

Table 4. Summary of stages of applying capability, opportunity, motivation, and behavior framework to the focus groups and interview data to identify nutrition, physical activity, and weight management intervention functions and features.

COM-B[a] and themes	User preferences	Intervention functions	Intervention features
Capability: psychological			
Misinformation	Simple expert advice; links to further information	Education; tailored feedback on performance	Mobile food record; dietary goals
Capability: physical			
Environmental support	Dietary self-monitoring tools is easy, quick, and subtle	Training: how to use the mobile food record app	Mobile food record
Environmental support	PA monitor is easy to use and is visually appealing	Training: how to use the Fitbit charge 2	Fitbit Charge 2; tailored movement goals
Opportunity: physical			
Environmental support	Feedback provides visual information, basis their food choice	Education: link to Australian guidelines for diet and PA[b]	Tailored dietary feedback
Opportunity: social			
Social norms	Feedback is nonjudgmental, supportive tone	Supportive and friendly tone	Tailored feedback, tailored education
Motivation			
Confidence	Goals for diet, PA, and weight change are realistic goals	Modelling; personalized examples of how to improve their current behavior	Tailored feedback; Fitbit Charge 2; tailored movement goals

[a]COM-B: capability, opportunity, motivation, and behavior.

[b]PA: physical activity.

Discussion

Principal Findings

This study reveals a lack of knowledge and confidence about evidence-based weight management behaviors and susceptibility to misinformation about nutrition. The results also suggest that the BCTs of self-monitoring and feedback on performance are well suited to this group. This paper follows previous examples of combining theory and the experience of users and experts to develop digital interventions [47]. Our findings explored factors affecting weight management behavior, all of which could be mapped to the COM-B model. Participants expressed concerns about their capability; misinformation and opportunities; availability of alternatives and motivation or plans to change their behavior [45,47].

The capability theme was most notable, with lack of knowledge and misinformation being most prevalent. Despite awareness and positive attitudes toward the LiveLighter campaign, participants were still unaware of how to implement the messages from the campaign into their own lives. Several researchers have identified this as nutrition literacy, the ability to interpret and use nutrition information [65-67]. Misinformation about effective dietary strategies to manage their weight was evident; the energy content of alcohol was underestimated and potatoes were described as *fattening*. This highlights the need for more tailored and specific information that addresses participants' ability to understand and implement new behaviors. A recent review recommends interventions provide actionable feedback and information on where and when to perform the new behavior [68]. For example, using the mFR to assess a meal image (Figure 1) and then providing actionable feedback such as identifying sources of EDNP foods and providing suggestions for change.

Motivation to change eating habits and reduce alcohol intake was hindered by beliefs, and government guidelines on alcohol were considered unrealistic, a view found previously [69]. Similarly, participants were aware of the recommendations to eat 2 serves of fruit and 5 serves of vegetables, but did not believe this was necessary [70]. This seems to be compounded by limited or inaccurate knowledge of serving sizes and energy content of foods [71].

Ambivalence appeared to be a strong theme, with participants describing their competing desires. Although most participants admitted to actively try to restrict their intake of junk food, they also revealed that junk foods were associated with happiness and *time off*. Similarly, eating out was viewed as a major barrier to weight management because of excess portion sizes and lack of healthy options. However, it was also described as a compulsory, normal part of our culture. These views are supported by a previous study, which highlighted that people felt pressured to eat junk food to participate socially and avoid criticism from their peers [72]. In addition, behavioral science has identified that our habits are the salient drivers of behavior, rather than motivation or intention to change [73]. This suggests that interventions should focus on developing habitual changes and creating healthy options as the default choice [73].

Social factors appear to be both motivators and barriers to healthy eating. Although most people agree that eating a healthy diet is what *should* be done, the social context of eating was associated with alcohol and junk food. Qualitative studies with Australian men aged 18 to 25 years reported that healthy eating was seen as incongruent with the masculine stereotype [44].

Our study confirmed that this view also exists among older men and women.

Strengths and Limitations

A major strength of the study was the large sample of participants living with overweight and obesity and the inclusion of male and female participants. Although men are more likely to have a poor diet, carry excess weight, and experience weight-related disease, they are underrepresented in weight loss interventions. As a result, this study sought to recruit male participants to ensure that their views and preferences were represented in the development of the intervention. The transcripts were coded by gender. However, no apparent differences in preferences or views were found between men and women. Further studies with this cohort would benefit from exploring specific gender differences in the experience of weight loss behaviors and gender preferences for digital intervention features.

This research explored digital technology on several platforms, including the mFR app, wearable PA trackers, email, and text messaging. This made it difficult to apply a single framework exclusively for each of these elements to the intervention. As a result, Australian guidelines for weight management, diet, and physical activity informed target behaviors, rather than a theory-based process such as the behavior change wheel, intervention mapping, or the Integrate Design Assess and Share framework [45,74,75].

The use of self-reported height and weight measures was a limitation of the study and may have led to inaccuracies in the reported BMI. At screening, volunteers were excluded if their BMI was <25. However, when asked at the focus groups to self-report their height and weight, 10 participants reported a BMI of <25. Therefore, the views expressed in these focus groups may not entirely reflect those living with overweight and obesity. Study participants primarily registered on the web via the LiveLighter website and were likely to already have an interest in digital weight management. Their preferences may differ from those who have not previously attempted to seek weight loss information online.

A limitation of the study is that the predetermined intention to develop a scalable digital behavior change intervention could have likely restricted the themes that emerged from the data. The purpose of this focus group study was to explore which intervention features would be acceptable and feasible to assist participants' behavior change. The study explored potential barriers and benefits of using technology, rather than the wider context of the lived experience of participants in relation to their weight issues, such as social support. This is a limitation of the study, and further in-depth research is needed to explore this issue. The strength of this research is that it explored participants' opinions on a variety of relatively accessible technological devices to gauge their suitability for intervention. Theoretically, these PA trackers are effective and ideal for hard-to-reach groups. Although they provide objective feedback, little is known about users' experiences or preferences regarding the use of these tools for self-monitoring purposes. Another limitation was that the discussion regarding social support was limited to identifying the desired frequency of researcher contact. This script focused specifically on the supportive features of the intervention and did not explore other social support as they were outside the scope of the study. A further limitation was that the pace at which this intervention was developed and evaluated was protracted in comparison with commercial interventions [76]. Recent industry and academic partnerships have demonstrated the potential to produce high-quality digital interventions at a commercial pace [77].

Future Directions

Dichotomous thinking about food and activity can impede efforts to make healthy lifestyle changes [78]. A more flexible and nonjudgmental approach can lead to better behavior change and reduce dietary restraint when supporting the psychological well-being of participants [39,79]. This intervention aims to adopt this flexible and nonjudgmental tone. The effectiveness of these strategies will be evaluated in a randomized controlled trial and exit interviews. Future studies will assess the relationship between behavior change and intervention features, consistent with guidelines for developing digital health interventions [49].

Conclusions

The ToDAy study was developed using a person-centered approach and behavior change theory. Focus groups and interviews were undertaken to explore user capability, opportunity, and motivation to perform the targeted behaviors for weight loss. The study revealed a lack of knowledge, confidence, and susceptibility to misinformation about evidence-based weight management behaviors. The findings suggest that a digital weight management intervention using mobile food records and activity trackers to inform tailored feedback may be an acceptable, feasible, and engaging strategy.

Acknowledgments

The authors would like to thank all the participants who generously gave up their time to participate in this study and Anne Finch, Iain Stephen Pratt (Cancer Council WA), and Joel Howell for their role as cofacilitators. CS is a PhD student supported by an Australian Government Research Training Program Scholarship. Funding for the ToDAy study was provided by a Healthway Health Promotion Research Grant and the East Metropolitan Health Service. The mFR app is funded by National Institute of Health, National Cancer Institute (1U01CA130784-01), and National Institute of Health, National Institute of Diabetes and Digestive and Kidney Diseases (1R01-DK073711-01A1, 2R56DK073711-04). The sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish the results. The term "LiveLighter" is a registered trademark. The LiveLighter campaign is funded by the Western Australian

Department of Health and at the time of this study, delivered by the National Heart Foundation (WA Division), and currently delivered by Cancer Council Western Australia.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Script for interviews and focus groups.
[DOCX File , 15 KB - mhealth_v8i9e17919_app1.docx]

References

1. Profiles of Health, Australia, 2011-13. Australian Bureau of Statistics, Australian Government. 2011. URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4338.0~2011-13~Main%20Features~Overweight%20and%20obesity~10007 [accessed 2018-11-14] [WebCite Cache ID 73un7bvI2]
2. Jaacks LM, Vandevijvere S, Pan A, McGowan CJ, Wallace C, Imamura F, et al. The obesity transition: stages of the global epidemic. Lancet Diabetes Endocrinol 2019 Mar;7(3):231-240 [FREE Full text] [doi: 10.1016/S2213-8587(19)30026-9] [Medline: 30704950]
3. Australian Health Survey: Nutrition First Results - Foods and Nutrients, 2011-12. Australian Bureau of Statistics, Australian Government. 2011. URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4364.0.55.007Main+Features12011-12 [accessed 2018-11-14] [WebCite Cache ID 73uoVjWZb]
4. Morley B, Niven P, Dixon H, Swanson M, Szybiak M, Shilton T, et al. Population-based evaluation of the 'LiveLighter' healthy weight and lifestyle mass media campaign. Health Educ Res 2016 Apr;31(2):121-135 [FREE Full text] [doi: 10.1093/her/cyw009] [Medline: 26956039]
5. Rhodes RE, Pfaeffli LA. Mediators of physical activity behaviour change among adult non-clinical populations: a review update. Int J Behav Nutr Phys Act 2010 May 11;7:37 [FREE Full text] [doi: 10.1186/1479-5868-7-37] [Medline: 20459781]
6. Hruby A, Hu FB. The epidemiology of obesity: a big picture. Pharmacoeconomics 2015 Jul;33(7):673-689 [FREE Full text] [doi: 10.1007/s40273-014-0243-x] [Medline: 25471927]
7. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]
8. LiveLighter. 2020. URL: https://livelighter.com.au/ [accessed 2018-08-29] [WebCite Cache ID 721f8bzFZ]
9. Morley B, Niven P, Dixon H, Swanson M, Szybiak M, Shilton T, et al. Association of the LiveLighter mass media campaign with consumption of sugar-sweetened beverages: cohort study. Health Promot J Austr 2019 Dec;30(Suppl 1):34-42 [FREE Full text] [doi: 10.1002/hpja.244] [Medline: 30903631]
10. Morley B, Niven PH, Dixon HG, Swanson MG, McAleese AB, Wakefield MA. Controlled cohort evaluation of the mass media campaign's impact on adults' reported consumption of sugar-sweetened beverages. BMJ Open 2018 Apr 25;8(4):e019574 [FREE Full text] [doi: 10.1136/bmjopen-2017-019574] [Medline: 29695387]
11. Browne J, MacDonald C, Egan M, Delbridge R, McAleese A, Morley B, et al. You wouldn't eat 16 teaspoons of sugar-so why drink it? Aboriginal and Torres Strait Islander responses to the LiveLighter sugary drink campaign. Health Promot J Austr 2019 Apr;30(2):212-218. [doi: 10.1002/hpja.196] [Medline: 30144362]
12. Kreuter MF, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27(Suppl 3):S227-S232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]
13. Brug J, Glanz K, van Assema P, Kok G, van Breukelen GJ. The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. Health Educ Behav 1998 Aug;25(4):517-531. [doi: 10.1177/109019819802500409] [Medline: 9690108]
14. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
15. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
16. Krebs P, Prochaska JO, 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]
17. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
18. Clinical Practice Guidelines for the Management of Overweight and Obesity. The National Health and Medical Research Council (NHMRC). URL: https://www.nhmrc.gov.au/about-us/publications/clinical-practice-guidelines-management-overweight-and-obesity [accessed 2019-02-07] [WebCite Cache ID 760BJX4o0]

19. Pirotta S, Joham A, Hochberg L, Moran L, Lim S, Hindle A, et al. Strategies to reduce attrition in weight loss interventions: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1400-1412. [doi: 10.1111/obr.12914] [Medline: 31347759]
20. Rollo M, Burrows T, Vincze LJ, Harvey J, Collins CE, Hutchesson MJ. Cost evaluation of providing evidence-based dietetic services for weight management in adults: in-person versus eHealth delivery. Nutr Diet 2018 Feb;75(1):35-43. [doi: 10.1111/1747-0080.12335] [Medline: 29411491]
21. Walthouwer MJL, Oenema A, Lechner L, de Vries H. Use and effectiveness of a video- and text-driven web-based computer-tailored intervention: randomized controlled trial. J Med Internet Res 2015 Sep 25;17(9):e222 [FREE Full text] [doi: 10.2196/jmir.4496] [Medline: 26408488]
22. Springvloet L, Lechner L, de Vries H, Candel MJ, Oenema A. Short- and medium-term efficacy of a web-based computer-tailored nutrition education intervention for adults including cognitive and environmental feedback: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e23 [FREE Full text] [doi: 10.2196/jmir.3837] [Medline: 25599828]
23. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
24. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
25. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
26. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
27. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What Matters in Weight Loss? An In-Depth Analysis of Self-Monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
28. Cleary M, Escott P, Horsfall J, Walter G, Jackson D. Qualitative research: the optimal scholarly means of understanding the patient experience. Issues Ment Health Nurs 2014 Nov;35(11):902-904. [doi: 10.3109/01612840.2014.965619] [Medline: 25353304]
29. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. Br Med J 2009 Sep 10;339:b3496 [FREE Full text] [doi: 10.1136/bmj.b3496] [Medline: 19744976]
30. Lozano-Sufrategui L. 'Sorry mate, you're probably a bit too fat to be able to do any of these': men's experiences of weight stigma. Int J Men's Health 2016;15(1):4-23 [FREE Full text]
31. Pagoto SL, Schneider KL, Oleski JL, Luciani JM, Bodenlos JS, Whited MC. Male inclusion in randomized controlled trials of lifestyle weight loss interventions. Obesity (Silver Spring) 2012 Jun;20(6):1234-1239 [FREE Full text] [doi: 10.1038/oby.2011.140] [Medline: 21633403]
32. Donnachie C, Wyke S, Mutrie N, Hunt K. 'It's like a personal motivator that you carried around wi' you': utilising self-determination theory to understand men's experiences of using pedometers to increase physical activity in a weight management programme. Int J Behav Nutr Phys Act 2017 May 5;14(1):61 [FREE Full text] [doi: 10.1186/s12966-017-0505-z] [Medline: 28476118]
33. Quested E, Kwasnicka D, Thøgersen-Ntoumani C, Gucciardi DF, Kerr DA, Hunt K, et al. Protocol for a gender-sensitised weight loss and healthy living programme for overweight and obese men delivered in Australian football league settings (Aussie-FIT): a feasibility and pilot randomised controlled trial. BMJ Open 2018 Oct 17;8(10):e022663. [doi: 10.1136/bmjopen-2018-022663] [Medline: 30337315]
34. Robertson C, Archibald D, Avenell A, Douglas F, Hoddinott P, van Teijlingen E, et al. Systematic reviews of and integrated report on the quantitative, qualitative and economic evidence base for the management of obesity in men. Health Technol Assess 2014 May;18(35):v-vi, xxiii [FREE Full text] [doi: 10.3310/hta18350] [Medline: 24857516]
35. Halse RE, Shoneye CL, Pollard CM, Jancey J, Scott JA, Pratt IS, et al. Improving nutrition and activity behaviors using digital technology and tailored feedback: protocol for the LiveLighter tailored diet and activity (today) randomized controlled trial. JMIR Res Protoc 2019 Feb 25;8(2):e12782 [FREE Full text] [doi: 10.2196/12782] [Medline: 30801257]
36. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]
37. 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]
38. Kerr DA, Harray AJ, Pollard CM, Dhaliwal SS, Delp EJ, Howat PA, et al. The connecting health and technology study: a 6-month randomized controlled trial to improve nutrition behaviours using a mobile food record and text messaging support in young adults. Int J Behav Nutr Phys Act 2016 Apr 21;13:52 [FREE Full text] [doi: 10.1186/s12966-016-0376-8] [Medline: 27098449]

39. Pollard CM, Howat PA, Pratt IS, Boushey CJ, Delp EJ, Kerr DA. Preferred tone of nutrition text messages for young adults: focus group testing. JMIR Mhealth Uhealth 2016 Jan 19;4(1):e1 [FREE Full text] [doi: 10.2196/mhealth.4764] [Medline: 26787115]
40. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
41. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
42. Educator Guide. Eat for Health. 2013. URL: https://www.eatforhealth.gov.au/sites/default/files/files/the_guidelines/n55b_eat_for_health_educators_guide.pdf [accessed 2019-02-07] [WebCite Cache ID 760D7SFfa]
43. Australia's Physical Activity and Sedentary Behaviour Guidelines and the Australian 24-Hour Movement Guidelines. Australian Government Department of Health. 2017. URL: http://www.health.gov.au/internet/main/publishing.nsf/content/health-pubhlth-strateg-phys-act-guidelines#apaadult [accessed 2020-02-07] [WebCite Cache ID 760EH2cX9]
44. Ashton L, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Young men's preferences for design and delivery of physical activity and nutrition interventions: a mixed-methods study. Am J Mens Health 2017 Sep;11(5):1588-1599 [FREE Full text] [doi: 10.1177/1557988317714141] [Medline: 28675118]
45. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
46. Miller CL. Integrating Consumer Engagement in Health and Medical Research – An Australian Framework. Health Research Policy and Systems. 2017. URL: https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-017-0171-2 [accessed 2020-07-30]
47. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR Mhealth Uhealth 2015 Jun 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
48. White BK, Martin A, White JA, Burns SK, Maycock BR, Giglia RC, et al. Theory-based design and development of a socially connected, gamified mobile app for men about breastfeeding (milk man). JMIR Mhealth Uhealth 2016 Jun 27;4(2):e81 [FREE Full text] [doi: 10.2196/mhealth.5652] [Medline: 27349756]
49. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
50. 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]
51. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva, Switzerland: World Health Organization; 2020.
52. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
53. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
54. Australian Alcohol Guidelines. The National Health and Medical Research Council (NHMRC). URL: https://alcoholthinkagain.com.au/alcohol-your-health/online-tools/australian-alcohol-guidelines/
55. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
56. Mummah SA, King AC, Gardner CD, Sutton S. Iterative development of vegethon: a theory-based mobile app intervention to increase vegetable consumption. Int J Behav Nutr Phys Act 2016 Aug 8;13:90 [FREE Full text] [doi: 10.1186/s12966-016-0400-z] [Medline: 27501724]
57. Boushey C, Spoden M, Delp E, Zhu F, Bosch M, Ahmad Z, et al. Reported energy intake accuracy compared to doubly labeled water and usability of the mobile food record among community dwelling adults. Nutrients 2017 Mar 22;9(3):312 [FREE Full text] [doi: 10.3390/nu9030312] [Medline: 28327502]
58. Ahmad Z, Kerr DA, Bosch M, Boushey CJ, Delp EJ, Khanna N, et al. A mobile food record for integrated dietary assessment. MADiMa16 (2016) 2016 Oct;2016:53-62 [FREE Full text] [doi: 10.1145/2986035.2986038] [Medline: 28691119]
59. Zhu F. An Image Analysis System for Dietary Assessment and Evaluation. In: IEEE International Conference on Image Processing. 2010 Presented at: ICIP'10; September 26-29, 2010; Hong Kong, China. [doi: 10.1109/icip.2010.5650848]

60. Shoneye C, Johnson F, Steptoe A, Wardle J. A qualitative analysis of black and white British women's attitudes to weight and weight control. J Hum Nutr Diet 2011 Dec;24(6):536-542. [doi: 10.1111/j.1365-277X.2011.01198.x] [Medline: 21838745]
61. Russell RD, Black LJ, Sherriff JL, Begley A. Dietary responses to a multiple sclerosis diagnosis: a qualitative study. Eur J Clin Nutr 2019 Apr;73(4):601-608. [doi: 10.1038/s41430-018-0252-5] [Medline: 29941913]
62. Begley A, Ringrose K, Giglia R, Scott J. Mothers' understanding of infant feeding guidelines and their associated practices: a qualitative analysis. Int J Environ Res Public Health 2019 Mar 29;16(7):1141 [FREE Full text] [doi: 10.3390/ijerph16071141] [Medline: 30934967]
63. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
64. Braun V, Victoria C, Nikki H, Gareth T. Thematic analysis. Res Method Health Soc Sci 2020;-:843-860. [doi: 10.1007/978-981-10-5251-4_103]
65. Velardo S. The nuances of health literacy, nutrition literacy, and food literacy. J Nutr Educ Behav 2015;47(4):385-9.e1. [doi: 10.1016/j.jneb.2015.04.328] [Medline: 26026651]
66. Silk KJ, Sherry J, Winn B, Keesecker N, Horodynski MA, Sayir A. Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities. J Nutr Educ Behav 2008;40(1):3-10. [doi: 10.1016/j.jneb.2007.08.012] [Medline: 18174098]
67. Gibbs H, Chapman-Novakofski K. Exploring nutrition literacy: attention to assessment and the skills clients need. Health 2012;4(3):120-124. [doi: 10.4236/health.2012.43019]
68. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
69. Harrison L, Kelly P, Lindsay J, Advocat J, Hickey C. 'I don't know anyone that has two drinks a day': young people, alcohol and the government of pleasure. Health, Risk & Society 2011 Aug;13(5):469-486. [doi: 10.1080/13698575.2011.596190]
70. Carter OB, Pollard CM, Atkins JF, Marie Milliner J, Pratt IS. 'We're not told why--we're just told': qualitative reflections about the Western Australian Go for 2&5 fruit and vegetable campaign. Public Health Nutr 2011 Jun;14(6):982-988. [doi: 10.1017/S1368980010003381] [Medline: 21205404]
71. Kothe E, Mullan B. Perceptions of fruit and vegetable dietary guidelines among Australian young adults. Nutr Dietetics 2011;68(4):262-266. [doi: 10.1111/j.1747-0080.2011.01550.x]
72. Brantley PJ, Stewart DW, Myers VH, Matthews-Ewald MR, Ard JD, Coughlin JW, et al. Psychosocial predictors of weight regain in the weight loss maintenance trial. J Behav Med 2014 Dec;37(6):1155-1168 [FREE Full text] [doi: 10.1007/s10865-014-9565-6] [Medline: 24722826]
73. Mullan B, Novoradovskaya E. Habit mechanisms and behavioural complexity. In: The Psychology of Habit. New York, USA: Springer; 2018.
74. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
75. Brendryen H, Johansen A, Nesvåg S, Kok G, Duckert F. Constructing a theory- and evidence-based treatment rationale for complex ehealth interventions: development of an online alcohol intervention using an intervention mapping approach. JMIR Res Protoc 2013 Jan 23;2(1):e6 [FREE Full text] [doi: 10.2196/resprot.2371] [Medline: 23612478]
76. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
77. Hendrie GA, James-Martin G, Williams G, Brindal E, Whyte B, Crook A. The development of vegeze: smartphone app to increase vegetable consumption in Australian adults. JMIR Form Res 2019 Mar 27;3(1):e10731 [FREE Full text] [doi: 10.2196/10731] [Medline: 30916653]
78. Palascha A, van Kleef E, van Trijp HC. How does thinking in Black and White terms relate to eating behavior and weight regain? J Health Psychol 2015 May;20(5):638-648. [doi: 10.1177/1359105315573440] [Medline: 25903250]
79. Sairanen E, Lappalainen R, Lapveteläinen A, Tolvanen A, Karhunen L. Flexibility in weight management. Eat Behav 2014 Apr;15(2):218-224. [doi: 10.1016/j.eatbeh.2014.01.008] [Medline: 24854807]

Abbreviations

BCT: behavior change technique
BIT: behavioral intervention technology
COM-B: capability, opportunity, motivation, and behavior framework
EDNP: energy-dense nutrient-poor
ToDAy: Tailored Diet and Activity
PA: physical activity

SSB: sugar-sweetened beverage

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by L Goff, C Valle, A Videira-Silva; comments to author 10.04.20; revised version received 15.06.20; accepted 07.07.20; published 09.09.20.

Please cite as:
Shoneye CL, Mullan B, Begley A, Pollard CM, Jancey J, Kerr DA
Design and Development of a Digital Weight Management Intervention (ToDAy): Qualitative Study
JMIR Mhealth Uhealth 2020;8(9):e17919
URL: https://mhealth.jmir.org/2020/9/e17919
doi:10.2196/17919
PMID:32641284

Original Paper

Testing Wearable UV Sensors to Improve Sun Protection in Young Adults at an Outdoor Festival: Field Study

Caitlin Horsham[1], MSc; Jodie Antrobus[2], MSc; Catherine M Olsen[3], PhD; Helen Ford[1], MSc; David Abernethy[1]; Elke Hacker[1], PhD

[1]School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia

[2]Preventive Health Branch, Queensland Health, Queensland Government, Brisbane, Australia

[3]Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, Australia

Corresponding Author:
Elke Hacker, PhD
School of Public Health and Social Work
Queensland University of Technology
60 Musk Avenue
Kelvin Grove
Brisbane, 4059
Australia
Phone: 61 731389674
Email: elke.hacker@qut.edu.au

Abstract

Background: Australia and New Zealand have the highest skin cancer incidence rates worldwide, and sun exposure is the main risk factor for developing skin cancer. Sun exposure during childhood and adolescence is a critical factor in developing skin cancer later in life.

Objective: This study aims to test the effectiveness of wearable UV sensors to increase sun protection habits (SPH) and prevent sunburn in adolescents.

Methods: During the weeklong school leavers outdoor festival (November 2019) at the Gold Coast, Australia, registered attendees aged 15-19 years were recruited into the field study. Participants were provided with a wearable UV sensor and free sunscreen. The primary outcome was sun exposure practices using the SPH index. Secondary outcomes were self-reported sunburns, sunscreen use, and satisfaction with the wearable UV sensor.

Results: A total of 663 participants were enrolled in the study, and complete data were available for 188 participants (188/663, 28.4% response rate). Participants provided with a wearable UV sensor significantly improved their use of sunglasses (P=.004) and sunscreen use both on the face (P<.001) and on other parts of the body (P=.005). However, the use of long-sleeve shirts (P<.001) and the use of a hat (P<.001) decreased. During the study period, 31.4% (59/188) of the participants reported receiving one or more sunburns. Satisfaction with the wearable UV sensor was high, with 73.4% (138/188) of participants reporting the UV sensor was helpful to remind them to use sun protection.

Conclusions: Devices that target health behaviors when outdoors, such as wearable UV sensors, may improve use of sunscreen and sunglasses in adolescents.

(JMIR Mhealth Uhealth 2020;8(9):e21243) doi:10.2196/21243

KEYWORDS

melanoma; health promotion; public health; preventive medicine; sunlight; sunburn; adolescents

Introduction

Skin cancer is Australia's national cancer and was estimated to account for more cases diagnosed than all other cancers combined and costing over US $580 million to treat each year [1-3]. Melanoma is the deadliest form of skin cancer and the most commonly diagnosed cancer in young adults aged 15-29 years in Australia [4]. Sunlight or ultraviolet radiation (UVR) is the main risk factor for skin cancer. Childhood and adolescence are critical periods during which exposure to UVR contributes to skin cancer in later life [5]. The amount of sun exposure received during this period is high, with half of our

total lifetime sun exposure received within the first 20 years [6]. Sunburn remains highly prevalent in Australia in the younger age groups, and as little as one severe sunburn in childhood can double the risk of developing a melanoma before the age of 40 years [6,7]. Australian adolescents continue to report high-risk behaviors such as spending long periods of time in the sun and positive views of tanning [8]. These findings are concerning and highlight the importance of preventative strategies during adolescence or young adult years, which may have a substantial impact on health outcomes later in life [9].

Daily sunscreen use at the population level has been shown to prevent keratinocyte cancers and melanoma deaths and to reduce health care costs [10]. Although adolescents have been reported to be knowledgeable about the risks of UVR and skin cancer, this knowledge does not always equate with the use of sunscreen or other sun protection practices [11,12]. Since the 1980s, mass media campaigns have been circulated in Australia, the most well-known being Slip, Slop, Slap, to raise awareness of the importance of sun protection [13]. To further motivate this hard-to-reach adolescent population, recent strategies have included social media campaigns, and positive impacts have been reported for the Sun Mum, Dear Melanoma, and Pretty Shady campaigns [14,15]. Technology including apps [16-18], UVR monitors or dosimeters [19], and UV detection stickers [20] can target people at the actual point of behavior as a *cue to action* [21]. The mobile phone app, Solar Cell, provided personalized time until sunburn information to over 600 US residents, aged >18 years, which led to a significant increase in sun protection behaviors [16]. Our research showed that young adults carrying a UVR dosimeter, which alarmed when UVR threshold levels were reached, reduced their time unprotected and exposed to UVR on weekends [19].

Outdoor festivals and mass gatherings of adolescents may pose certain health risks [22]. School leaver festivals are common events across Australia and are often referred to as *schoolies celebrations* to mark the graduation of students from the secondary education system. The Gold Coast school leavers festival is ticketed and incorporates free outdoor music concerts and beach activities in a high UVR environment and has been operating for over 10 years with between 16,000 and 18,000 registered attendees each year [23]. During festivals, some youth may engage in risky health behaviors, such as alcohol consumption or using illicit drugs [24]. Sunburn is also common at outdoor festivals in Australia when people are exposed to UVR during peak hours for long periods of time, and low adherence to the use of sun protective clothing is frequently observed [25]. An observational study reported that 14% of event attendees wore long-sleeve shirts, 56% wore hats, and 83% wore sunglasses [25]. In Australia, festival and event organizers are responsible for providing a safe environment for all attendees and staff [26]. Implementation of harm minimization strategies, including effective preplanning and resource provisions, is necessary to produce a safe environment. At the school leavers festival on the Gold Coast, all attendees must wear a wristband to gain entry to events. The addition of a sun safety prompt to this wristband could provide tailored health information to each user as a call to action, with feedback provided when it has the most potential to be beneficial during UVR exposure [21].

This research describes the development of a wearable UV sensor and presents the findings of a field study testing if the device could improve sun protection behaviors and reduce sunburn among adolescents at a weeklong outdoor school leavers festival.

Methods

Field Study Participants and Setting

Participants in the field study were attending the weeklong school leavers outdoor festival at the Surfers Paradise beach on the Gold Coast (latitude 28^{o}S, 153^{o}E) during November 16-22, 2019 (spring) in Australia.

The field study was approved by the Human Research Ethics Committee of the Queensland University of Technology (QUT; #1900000435) and prospectively registered with the Australian and New Zealand clinical trials register (ACTRN12619000976189).

Wearable UV Sensor Development

The wearable UV sensor was developed by the researchers for the school leavers outdoor festival in collaboration with the event organizers. The UV sensor was fitted to the event wristband and was developed to change color when exposed to sunlight. The UV sensor component was developed using a UV-responsive photochromic dye and molded into a silicon slider to fit over the existing wristband fabric. The event wristband is designed not to be removed once worn, and the fabric material has a locking mechanism that requires the wristband to be cut for removal. On the UV sensor, the slogan *be safe and watch your mates* was added as a reminder to avoid risky situations during the festival. The measurement accuracy of the silicon slider was determined using a UV intensity meter (Solar Light Co, Model PMA2100) fitted with a digital sensor (Solar Light Co, model PMA2101). Prototype and safety testing were undertaken in Brisbane (latitude 27^{o}S, 153^{o}E) during May (autumn) in Australia (Multimedia Appendix 1).

Field Study Recruitment and Treatment Regimen

At the start of the outdoor festival, school leavers attended a registration event where participants were recruited, completed a baseline questionnaire, and were provided with their choice of a free tube of sunscreen (Cancer Council SPF 50+, 75 mL or 110 mL) for use during the festival. All individuals attending the school leavers outdoor festival were provided with the wearable UV sensor regardless of their participation in the study. The wearable UV sensor was required to be worn by the festival attendees as a wristband for the duration of the festival to gain entry to the festival events. During the registration process, event staff secure the wristband to the attendees' wrist. If the wristband is removed during the festival, participants are required to obtain a replacement wristband to gain entry to events. Posters displayed at the registration event described how the wearable UV sensors functioned (Multimedia Appendix 1). A follow-up web-based survey was emailed 7 days after the participant attended the outdoor festival, which asked about

their sun protection behaviors during the festival (November 17-22, 2019). The follow-up survey did not collect sun protection behavior data for the first day of the festival, when participants enrolled in the study and collected their intervention device. Data collection was managed using REDCap electronic data capture tools hosted at the QUT, and surveys are shown in Multimedia Appendix 2 [27]. Participants received an initial email initiation to complete the follow-up survey on the web, followed by two reminders.

Weather Measurements During the Field Study

UVR data were recorded using a UV-Biometer model 501 (Solar Light Co), and data were displayed using the UV index scale. The standard erythemal dose (SED) was also calculated using daily summaries and hourly observations recorded at 10 AM and at noon. The UVR data were captured by the Australian Radiation Protection and Nuclear Safety Agency detector (Gold Coast, latitude 28^{o}S, 153^{o}E).

The proportion of cloud cover in the sky above the Surfers Paradise beach was recorded twice a day in the morning between 8 AM and 10 AM and in the afternoon between 2 PM and 4 PM. Images of the sky above the beach were captured using a fixed camera maintained by Coastalwatch [28]. The proportion of cloud cover in each image was counted using ImageJ software [29], as described previously [20]. All field trial image analysis and quantification procedures were performed blind to the image ID.

Temperature data were recorded in degrees Celsius, and data were reported for the daily minimum and maximum as well as observations at 9 AM and 3 PM each day. The temperature data were captured by the Bureau of Meteorology weather station (040764; Gold Coast Seaway, latitude 28^{o}S, 153^{o}E).

Outcome Measures

The primary outcome measure was the sun protection habits (SPH) index described by Glanz et al [30] and updated by Heckman et al [31] for young adults, which queries the frequency of 7 sun protective habits that are used when outdoors using a 4-point Likert scale (1=never or rarely, 2=sometimes, 3=usually, and 4=always), which are averaged to derive the score, including wearing a shirt with sleeves, wearing a hat, wearing sunglasses, wearing sunscreen with a sun protection factor (SPF) 15 or higher on the face, wearing sunscreen with an SPF 15 or higher on other parts of the body, staying in the shade, and limiting time in the sun during midday hours. Secondary outcomes were self-reported, including the number of sunburns; sunburn intensity (mild, moderate, or severe) and location of sunburn; sunscreen use, including the frequency of daily application; and satisfaction with the wearable UV sensor. In the follow-up survey, participants completed an open-answer question on their comments or suggestions about the study.

Statistical Analysis

The Pearson chi-squared and/or Fisher exact test was used to detect the statistical significance in the difference between the groups who completed the follow-up survey and those who did not. For the participants who completed the follow-up survey, the Wilcoxon matched-pairs signed-rank test was used to examine the differences in sun protection behaviors at baseline and at follow-up. Changes in SPH were measured, and participants were grouped by comparing an individual's baseline value with their follow-up value and allocating them to either the improved, decreased, or no change groups. Analyses were performed using SAS statistical software package (SAS Institute).

Inductive thematic analysis was used to group open-ended answers into themes by 2 researchers (CH and DA).

Results

Wearable UV Sensor Observational Testing

The wearable UV sensor comprises a UV-responsive photochromic dye that is white when not in sunlight and turns purple in sunlight, indicating that sun protection is required (Figure 1). The threshold for the color change was tested using a graded series of UVR intensities. The color change was observed in part shade when the UVR level was low, 0.21 uW/cm^2, and no color change was observed when the indicator was in full shade, 0.0 uW/cm^2. When the sensor was placed in the sun (UVR=5.27 uW/cm^2), the sensor immediately changed color to purple (Multimedia Appendix 1). The photochromic dye in the wearable UV sensor was shown to be responsive to low-level UVR intensity, demonstrating that it was an adequate indicator for use in the field study.

Figure 1. Wearable UV sensor. The wearable UV sensor is white (left image) which indicates no UV exposure. The UV indicator is purple (right image) demonstrating exposure to UV radiation.

Field Study Participant Characteristics

During the first day of the festival, 663 volunteers were enrolled and completed the baseline survey. The evaluation survey was completed by 188 (28.4%) participants (Table 1; Figure 2). There were demographic differences between those who completed the study (n=188) and those who did not (n=475), with females and people with brown hair more likely to complete the study (Table 1).

Of the 188 participants who completed the study, most participants had very fair or fair skin (114/188, 60.6%), were women (145/188, 77.1%), and their age ranged from 15 to 19 years (Table 1).

Figure 2. During recruitment participants completed either a written or web-based baseline survey, and some email addresses were not completed correctly; invalid email address=email bounced back and the follow-up survey was not received by the participant.

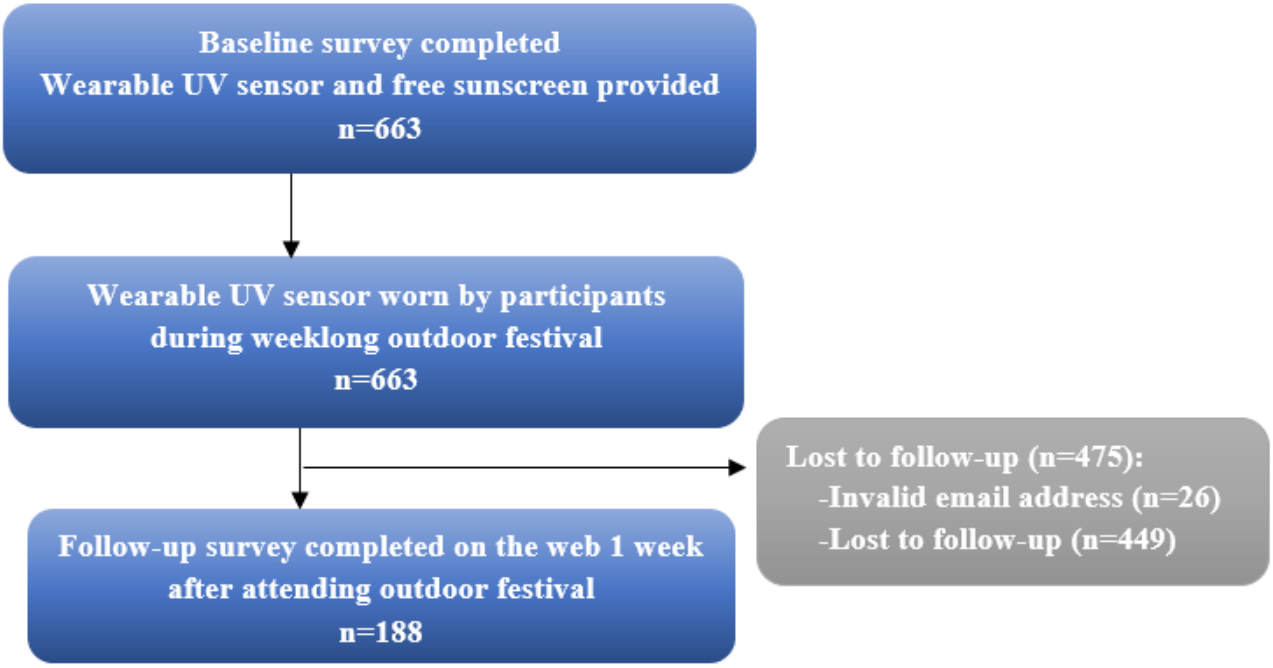


Table 1. Field study participant characteristics.

Characteristics	Baseline (n=663)	No evaluation; (n=475)	Evaluation (n=188)	*P* value
Age (years), range	15-19	17-19	15-18	.12
Gender, n (%)				.006
Female	452 (68.2)	307 (64.6)	145 (77.1)	
Male	208 (31.4)	165 (34.7)	43 (22.9)	
Other	3 (0.5)	3 (0.6)	0 (0)	
Skin color, n (%)				.25
Very fair or fair	379 (57.2)	265 (55.8)	114 (60.6)	
Medium	186 (28.1)	139 (74.0)	47 (25.0)	
Olive or brown	97 (14.6)	71 (15.0)	26 (13.8)	
Missing	1 (0.2)	0 (0)	1 (0.5)	
Hair color, n (%)				.008
Red (including auburn)	54 (8.1)	40 (8.4)	14 (7.4)	
Fair or blonde (including white)	154 (23.2)	126 (26.5)	28 (14.9)	
Light brown, mouse brown, or dark brown	418 (63.1)	282 (59.4)	136 (72.3)	
Black	36 (5.4)	27 (5.7)	9 (4.8)	
Missing	1 (0.2)	0 (0)	1 (0.5)	
Skin burn in strong summer sun for 30 min without protection?, n (%)				.15
My skin would not burn at all	99 (14.9)	78 (16.4)	21 (11.2)	
My skin would burn lightly	286 (43.1)	203 (42.7)	83 (44.2)	
My skin would burn moderately	228 (34.4)	163 (34.3)	65 (34.6)	
My skin would burn severely	49 (7.4)	31 (6.5)	18 (9.6)	
Missing	1 (0.2)		1 (0.5)	
Skin tan in strong sun without protection?, n (%)				.17
My skin would not tan	122 (18.4)	79 (16.6)	43 (22.9)	
My skin would tan lightly	224 (33.8)	164 (34.5)	60 (31.9)	
My skin would tan moderately	233 (35.1)	175 (36.8)	58 (30.9)	
My skin would tan deeply	84 (12.7)	57 (12.0)	27 (14.4)	
Have you made an attempt to get a suntan in the last 12 months?, n (%)				.49
Yes	435 (65.6)	317 (66.7)	118 (62.8)	
No	227 (34.2)	157 (33.1)	70 (37.2)	
Missing	1 (0.2)	1 (0.2)	0 (0)	
During the past 12 months, how many times did you get sunburnt?, n (%)				.31
Never	43 (6.5)	34 (7.2)	9 (4.8)	
Once	138 (20.8)	102 (21.5)	36 (19.1)	
2-5 times	349 (52.6)	239 (50.3)	110 (58.5)	
≥6 times	97 (14.6)	71 (15.0)	26 (13.8)	
Do not know or unsure	36 (5.5)	29 (6.1)	7 (3.7)	

Sun Protection Items Brought to the Festival by Participants

Sun protection items were common among all participants (n=663), with 76.6% (508/663) bringing a hat, 76.0% (504/663) bringing sunglasses, and 66.0% (438/663) bringing sunscreen, whereas only 41.6% (276/663) brought a long-sleeve shirt and 12.2% (81/663) brought a beach umbrella (Multimedia Appendix 2).

Weather Conditions During the Field Study

The UVR exposure level was consistently high, requiring sun protection each day during the weeklong outdoor festival with daily SEDs ranging from 36 SEDs to 54 SEDs (Table 2). The UV index level was >3 for over 6 hours each day of the festival, the cloud cover was consistently clear throughout the week, and there was no rainfall recorded (Table 2). The average daily maximum temperature was 28°C (range 26.1-30.1) during the field study (Table 2).

Table 2. Weather conditions during the field study.

Date	Start of day when UVI[a] >3, AM	End of day when UVI <3, PM	Temperature, °C				UV daily[b] dose	UV 10-11 AM[c] dose	UV noon -1 PM[d] dose	Cloud cover	Rain (mm)
			Minimum	Maximum	9 AM	3 PM	SEDs[e]	SEDs	SEDs		
November 16, 2019	8	3	21.5	30.1	28.7	26.1	47	8	8	Nil	0
November 17, 2019	8	2	19.9	27.6	24.4	23.9	42	7	8	Nil	0
November 18, 2019	8:30	3:30	19.5	26.1	23.5	23.9	54	9	9	Nil	0
November 19, 2019	8:30	2:30	17.7	29.2	25.0	25.6	36	7	6	Nil	0
November 20, 2019	8:30	2:30	22.1	29.5	27.0	26.9	43	7	7	Nil	0
November 21, 2019	8:30	3	21.7	27.4	25.7	25.3	47	7	8	Nil	0
November 22, 2019	8:30	3	19.9	27.0	24.8	25.5	42	7	6	Nil	0

[a]UVI: ultraviolet index.

[b]Daily dose calculated from 6 AM-4 PM.

[c]Morning dose calculated from 10 AM-11 AM.

[d]Midday dose calculated from noon-1 PM.

[e]SEDs: standard erythemal dose.

SPH Index

At baseline (n=188), the mean SPH index value was 2.31 (SE 0.04) and did not change at follow-up (+0.03; Table 3). Some individual SPH items improved significantly at follow-up compared with baseline, including use of sunglasses (+0.21; *P*=.004), sunscreen use on the face (+0.36; *P*<.001), and sunscreen use on the body (+0.22; *P*=.005). The use of long-sleeve shirts (−0.31; *P*<.001) and use of a hat (−0.30; *P*<.001) both decreased significantly at follow-up.

Over 41.4% (78/188) of participants improved their Likert scale level from baseline for their use of sunscreen applied to the face, whereas 17.0% (32/188) of the participants decreased their usage and 41.5% (78/188) of the participants had no change at follow-up. Sunscreen applied to the body also improved, with 38.3% (72/188) of participants improving their Likert scale score, whereas 20.7% (39/188) of the participants decreased their usage and 40.9% (77/188) of the participants had no change at follow-up.

Table 3. Sun protection habits index.

Items	Baseline (n=188), mean (SE)	Follow-up (n=188), mean (SE)	Change	*P* value[a]
SPH[b] index (below items combined)	2.31 (0.04)	2.34 (0.03)	+0.03	No change
Wear a shirt with long sleeves	1.66 (0.05)	1.35 (0.05)	−0.31	<.001
Wear sunglasses	2.30 (0.07)	2.51 (0.08)	+0.21	.004
Stay in the shade	2.50 (0.03)	2.60 (0.05)	+0.10	.31
Limit your time in the sun during midday hours	2.40 (0.06)	2.30 (0.06)	−0.10	.48
Wear a hat	2.41 (0.07)	2.11 (0.07)	−0.30	<.001
Wear sunscreen with an SPF[c] ≥15 on your face	2.46 (0.06)	2.82 (0.07)	+0.36	<.001
Wear sunscreen with an SPF[c] ≥15 on other parts of your body	2.43 (0.06)	2.65 (0.06)	+0.22	.005

[a]Wilcoxon matched-pairs signed rank test.

[b]SPH: sun protection habits. 1=never or rarely, 2=sometimes, 3=usually, and 4=always.

[c]SPF: sun protection factor.

Sunburn, Sunscreen Usage, and Time Outdoors

The baseline survey included a question about sunburns during the week before the festival, and 36.7% (69/188) of the participants reported one or more sunburns. During the festival, 31.4% (59/188) of the participants reported being sunburnt (Table 4). During the 12 months before the festival, over 91.4% (172/188) of the participants reported one or more sunburns (Table 1).

During the outdoor festival, sunburns were most commonly reported on the shoulders (36/109, 33.0%), followed by the head or face (27/109, 24.8%), neck (19/109, 17.4%), and legs (10/109, 9.2%; Table 4). The majority of sunburns reported were of mild intensity (71/109, 65.1%), followed by moderate (32/109, 29.4%) and severe (4/109, 3.7%; Table 4).

Participants who reported seeking a deliberate suntan in the previous year were more likely to report a sunburn during the outdoor festival (75% vs 57%; *P*=.02; Multimedia Appendix 2). Participants who reported a sunburn were also more likely to have not worn the wearable UV sensor (9% vs 2%; *P*=.05; Multimedia Appendix 2).

Sunscreen use was commonly reported in the follow-up survey, with 88.3% (166/188) of participants applying sunscreen one or more times during the day at the outdoor festival (Table 4). Most participants who reported being sunburned also reported applying sunscreen (51/59, 86%), and just under half of those sunburnt reported reapplying sunscreen two or more times per day (27/59, 46%).

Most participants (115/188, 61.2%) spent 2 hours outside each day of the festival; 30.8% (58/188) of the participants spent 1-2 hours outdoors and only 7.9% (15/188) spent ≤1 hour outdoors each day.

Table 4. Sunburn, sun tanning, and sunscreen use.

Survey items	Baseline (N=188), n (%)	Follow-up (N=188), n (%)
Did you experience any sunburn?[a]		
Yes	69 (36.7)	59 (31.4)
No	119 (63.3)	129 (68.6)
Total number of sunburn events[b]	142	109
Location of sunburns		
Body locations below combined	142 (100)	109 (100)
Head or face	50 (35.2)	27 (24.8)
Neck	20 (14.1)	19 (17.4)
Shoulders	30 (21.1)	36 (33.0)
Back	10 (7.1)	4 (3.7)
Chest	1 (0.7)	0 (0)
Arms	14 (9.9)	6 (5.5)
Hands	3 (2.1)	1 (0.9)
Legs	7 (4.9)	10 (9.2)
Feet	7 (4.9)	5 (4.6)
Buttocks	0 (0)	1 (0.9)
Intensity of sunburns		
Mild (pink to light redness)	64 (45.1)	71 (65.1)
Moderate (red skin)	66 (46.5)	32 (29.4)
Severe (deep redness, blisters may develop)	5 (3.5)	4 (3.7)
Missing	7 (4.9)	2 (1.8)
Have you tried to get a suntan during the festival?		
Yes	N/A[c]	97 (51.6)
No	N/A	91 (48.4)
Number of sunscreen applications per day during the festival		
0	N/A	22 (11.7)
1	N/A	86 (45.7)
≥2	N/A	80 (42.6)

[a]Baseline data collection: During the past week, how many times did you get sunburnt? Follow-up data collection: "We would like to know if you experienced any sunburn during November 17th to November 22nd 2019?"

[b]Incidence of sunburnt body areas. Participants may have received multiple sunburn events.

[c]N/A: not applicable.

Satisfaction With the Wearable UV Sensor

Adherence to the intervention device was high, with 95.7% (180/188) of the participants wearing the wearable UV sensor during the outdoor festival (Multimedia Appendix 2). Over 73.4% (138/188) of the participants found the wearable UV sensor helpful to remind them to apply sunscreen, whereas 16.5% (31/188) reported that it was not helpful and 10.1% (19/188) were unsure. Over 83.5% (157/188) of the participants would like to have this product included on wristbands for future daytime outdoor festivals. On a scale from 1 (not at all satisfied) to 10 (extremely satisfied), participants' mean satisfaction score was high at 8.1 (range 2-10; Multimedia Appendix 2).

Open-ended questions were completed by 50.0% (94/188) of the participants. The themes discussed included helpful, reinforced behaviors, recommendations for improvements, lack of awareness, enjoyment, and lack of impact and appearance. A participant commented that the wristband and embedded UV sensor was fashionable "I thought it was a cool item that wasn't overly apparent, so it went with the things you were wearing." The description of each theme along with example comments from participants are shown in Multimedia Appendix 2.

Discussion

Principal Findings

In this study we developed a wearable sun safety device and investigated the impact of the device to improve adolescents' use of sun protection behaviors at an outdoor festival. We found that providing participants with a wearable UV sensor improved their use of individual SPH items, including use of sunscreen and sunglasses. There was no improvement in the combined SPH index as a result of participants decreasing their use of long-sleeve shirts, along with the use of a hat. There was no significant change in the use of shade or time spent outdoors, which is not surprising given the outdoor festival context. Only under one-third of the participants experienced a sunburn during the outdoor festival, with the shoulders most commonly sunburnt. Our results indicate that participants who had a history of deliberate suntanning were significantly more likely to receive a sunburn during the outdoor festival. Participants reported high satisfaction rates, with most participants reporting that they would like to have the wearable UV sensor available at future outdoor events.

Sun protection is a multifaceted behavior in adolescents, involving a wide range of factors. In this field study, participants who exhibited high-risk behaviors, such as deliberate tanning, were more likely to receive a sunburn. A previous study of 4150 adolescents aged 14-17 years showed this association between adolescent sunburn and a positive attitude toward tanning as well as high rates of tanning, with 48% reporting they liked to tan [32]. The desire for tanned skin has been reported as a significant barrier to using sun protection in adolescents, with a tan associated with providing a sense of confidence, achievement, attractiveness, and the ability to fit in with their peers [8]. A photoaging app, called Sunface, which predicts the effects of sun damage to the face (such as wrinkles or aging) was tested in adolescents and was found to be effective in changing tanning and sun safe behaviors [33]. During 2010-2014, the nationwide campaign *The dark side of tanning* targeted young Australians' attitudes toward tanning and was shown to be effective at reducing positive attitudes toward tanning [34]. In Australia's nationwide sun protection telephone survey, 60% of adolescents reported a preference to tan in 2003, which dropped to 38% in 2013, and remained unchanged in 2019 [35]. A review of interventions to reduce tanning has shown that many programs are designed to increase knowledge of the risks of tanning and that there is the potential to formulate new programs to incorporate relevant addiction science techniques, including the use of brief motivational and cognitive behavioral-based interventions [36]. Tanning behavior is complex and multifaceted, and interventions are needed to address this issue as a high proportion of adolescents are still partaking in this high-risk behavior.

In adolescents, clothing choices are mostly influenced by fashion trends rather than sun protection [8]. A survey of young female beachgoers aged 17-35 years in Australia found that their sun protection at the beach was influenced by being uncomfortable or unstylish and whether friends or peers approved of their sun safe behavior [37]. In our study, sun protective clothing such as long-sleeve shirts or hats did not improve during the festival, whereas the use of sunglasses did. The use of long-sleeved shirts may not be considered fashionable or practical in hot environments, and *hat hair* after using a hat, which can negatively alter a hairstyle, has previously been cited as a barrier to hat use in adolescents [8]. To effectively target sun protection behavior in youth culture, further interventions may need to focus on changing adolescents' perceptions of what is healthy and fashionable [38].

A key priority for skin cancer prevention technology should be to understand the factors that reduce the incidence of sunburn. In our study, 31% of the participants reported one or more sunburns during the festival, which is similar to previous estimates showing 26% of Australian adolescents aged 12-17 years reported being sunburnt on summer weekends [39]. Many participants in this study reporting sunburns also reported wearing and reapplying sunscreen, suggesting sunscreen was not applied sufficiently to provide adequate UV protection. The reasons that people may receive an unintended sunburn after sunscreen use has been explored using qualitative interviews, which showed adults overestimated the amount of time they could safely be exposed to the sun and not reapplying sunscreen often enough, especially during water-based activities [40]. This study was based on a beach environment that may involve swimming, and this requires more frequent sunscreen applications. The wearable UV sensor was worn by almost all participants for the duration of the study, highlighting the robustness of the device.

Another key barrier to sun protection reported by adolescents is forgetting to protect their skin [41,42]. This may explain why adolescents have good sun protection knowledge but still report high rates of sunburn. To reduce the impact of forgetfulness, interventions that act as a reminder to target behaviors at the actual point of need are required [21]. The wearable UV sensor developed in this study, which changes color in the sun notifying the user that sun protection is required, is an example of a real-time, wearable prompt for sun safety. Other technologies such as UV detection stickers have been shown to be useful prompts for the re-application of sunscreen. Our research previously tested UV detection stickers in 428 adults during an outdoor sporting event with the aim of improving the re-application of sunscreen [20]. We found that participants provided with a UV detection sticker were more likely to re-apply sunscreen than controls (80% vs 68%; P=.04); however, the stickers did not reduce sunburn rates [20]. The wearable UV sensors tested in this study notified the user when they were exposed to UVR, whereas UV detection stickers prompt the user when their sunscreen required re-application. These ecological momentary interventions are examples of reminders to influence behaviors within an environmental context, as tailoring the content of health messages based on individual characteristics can improve willingness to change [21].

Strengths and Limitations

The strengths of this study include recruiting a large sample of adolescents who wore a UV sensor for 1 week in a high UV setting. This target group is typically difficult to reach in public health research, and the findings of this study could be applied

to other high UV environments such as sporting events or summer festivals. Limitations of this study include self-reported outcome measures, including wearable UV sensor compliance as well as the low response rate for the follow-up survey, and no datasets were collected on long-term sustained behavior change. The wearable UV sensor developed in this study was designed to not be removed once worn and was required to be viewed each day for festival entry. Only 4.3% (8/188) of the participants in our study reported removing the UV sensor wristband and not wearing it during the outdoor festival, and we did not collect further data on if they obtained a replacement band. The Hawthorne effect, where study participants alter their behavior because of being observed, could be a potential factor for the increase in sunscreen use [43]. The extent to which the provision of free sunscreen contributed to behavior change compared with wearing the UV sensor remains unknown. However, in Australia, legislation states that all people who are exposed to sunlight at outdoor events should be able to re-apply sunscreen every 2 hours. Guidelines further stipulate that if security measures prevent sunscreen from being brought inside the venue, event organizers have a duty of care to provide easy access within the venue. In Australia, it is common for sunscreen to be promoted and supplied by event organizers at outdoor mass gathering as standard care.

Participants who completed the study were mainly young females, and the results may not be generalizable to other subgroups of the population. However, adolescents are commonly underrepresented in prevention studies, spend long periods of time in the sun, and have higher rates of sunburn than adults [44]. Selection bias was a further limitation as participants were recruited at this event using a sequential, convenience sample, and we did not use a random sampling method.

Conclusions

This study developed and tested the effectiveness of a wearable UV sensor to improve sun protection and decrease sunburn in adolescents at an outdoor festival. Provision of a wearable UV sensor and free sunscreen improved use of sunglasses and sunscreen in participants. The wearable UV sensors did not reduce sunburn rates, and those who reported a history of suntanning were more likely to be sunburnt. The wearable UV sensor technology resonated with adolescent participants in this study with high satisfaction rates, and participants found them to be a helpful reminder for sun protection.

Acknowledgments

The authors would like to thank all participants for their time and the project staff (Erin Potter, Juliana Mahon, Darshana Munasinghe, Melanie O'Meara, and Lisa Naumann) for their contributions to the study. EH was supported by a fellowship from the Advance Queensland Industry Research Fellowship at QUT. Queensland Health and the Department of Child Safety, Youth and Women donated the wearable UV sensors and sunscreen for the study. UVR measurements were collected by the Australian Radiation Protection and Nuclear Safety Agency UV network and provided to the research team. The sponsors of the study (Queensland Government Advance Queensland fund) had no role in the study design, collection, analysis, and interpretation of data; in the writing of this manuscript; and in the decision to submit the paper for publication. The corresponding author had full access to all data in the study and final responsibility for the decision to submit for publication.

Authors' Contributions

EH contributed to conceptualization; funding acquisition; investigation; methodology; project administration; resources; data curation; formal analysis; supervision; visualization; and writing, reviewing, and editing the manuscript. CH contributed to project administration; investigation; data curation; formal analysis; and writing, reviewing, and editing the manuscript. DA contributed to project administration, data curation, formal analysis, and reviewing and editing the manuscript. HF contributed to project administration and reviewing and editing the manuscript. CO contributed to formal analysis and reviewing and editing of the manuscript. JA contributed to the investigation and reviewing and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures.
[DOCX File , 1410 KB - mhealth_v8i9e21243_app1.docx]

Multimedia Appendix 2
Supplementary tables.
[DOCX File , 34 KB - mhealth_v8i9e21243_app2.docx]

References

1. Australian Institute of Health and Welfare (AIHW). Skin cancer in Australia. Cat no. CAN 96. Canberra, Australia: AIHW; 2016.

2. Fransen M, Karahalios A, Sharma N, English DR, Giles GG, Sinclair RD. Non-melanoma skin cancer in Australia. Med J Aust 2012 Nov 19;197(10):565-568. [doi: 10.5694/mja12.10654] [Medline: 23163687]
3. Elliott TM, Whiteman DC, Olsen CM, Gordon LG. Estimated healthcare costs of melanoma in Australia over 3 years post-diagnosis. Appl Health Econ Health Policy 2017 Dec;15(6):805-816. [doi: 10.1007/s40258-017-0341-y] [Medline: 28756584]
4. Australian Institute of Health and Welfare (AIHW). Cancer in adolescents and young adults in Australia. Cat. no. CAN 110. Canberra: AIHW 2018.
5. Whiteman DC, Whiteman CA, Green AC. Childhood sun exposure as a risk factor for melanoma: a systematic review of epidemiologic studies. Cancer Causes Control 2001 Jan;12(1):69-82. [doi: 10.1023/a:1008980919928] [Medline: 11227927]
6. Green AC, Wallingford SC, McBride P. Childhood exposure to ultraviolet radiation and harmful skin effects: epidemiological evidence. Prog Biophys Mol Biol 2011 Dec;107(3):349-355 [FREE Full text] [doi: 10.1016/j.pbiomolbio.2011.08.010] [Medline: 21907230]
7. Cust AE, Jenkins MA, Goumas C, Armstrong BK, Schmid H, Aitken JF, et al. Early-life sun exposure and risk of melanoma before age 40 years. Cancer Causes Control 2011 Jun;22(6):885-897. [doi: 10.1007/s10552-011-9762-3] [Medline: 21472378]
8. Potente S, Coppa K, Williams A, Engels R. Legally brown: using ethnographic methods to understand sun protection attitudes and behaviours among young Australians 'I didn't mean to get burnt--it just happened!'. Health Educ Res 2011 Mar;26(1):39-52. [doi: 10.1093/her/cyq066] [Medline: 21059798]
9. Oliveria SA, Saraiya M, Geller AC, Heneghan MK, Jorgensen C. Sun exposure and risk of melanoma. Arch Dis Child 2006 Mar;91(2):131-138 [FREE Full text] [doi: 10.1136/adc.2005.086918] [Medline: 16326797]
10. Gordon L, Olsen C, Whiteman DC, Elliott TM, Janda M, Green A. Prevention versus early detection for long-term control of melanoma and keratinocyte carcinomas: a cost-effectiveness modelling study. BMJ Open 2020 Feb 26;10(2):e034388 [FREE Full text] [doi: 10.1136/bmjopen-2019-034388] [Medline: 32107270]
11. Heckman CJ, Manne SL, Kloss JD, Bass SB, Collins B, Lessin SR. Beliefs and intentions for skin protection and UV exposure in young adults. Am J Health Behav 2011 Nov;35(6):699-711 [FREE Full text] [Medline: 22251761]
12. Volkov A, Dobbinson S, Wakefield M, Slevin T. Seven-year trends in sun protection and sunburn among Australian adolescents and adults. Aust N Z J Public Health 2013 Mar;37(1):63-69. [doi: 10.1111/1753-6405.12012] [Medline: 23379808]
13. Iannacone MR, Green AC. Towards skin cancer prevention and early detection: evolution of skin cancer awareness campaigns in Australia. Melanoma Manag 2014 Aug;1(1):75-84 [FREE Full text] [doi: 10.2217/mmt.14.6] [Medline: 30190812]
14. Department of Health Annual Report 2014–15. Queensland Health. 2015. URL: https://www.health.qld.gov.au/__data/assets/pdf_file/0021/439050/annual-report-14-15.pdf [accessed 2020-04-20]
15. Nsw Skin Cancer Prevention Strategy 2012–15: Final Draft Evaluation Report. Cancer Institute NSW: Lessening the Impact of Cancer. URL: https://www.cancer.nsw.gov.au/getattachment/how-we-help/cancer-prevention/skin-cancer-prevention/nsw-skin-cancer-strategy/NSW-Skin-Cancer-Prevention-Strategy-2012-15-Evaluation-Report.PDF?lang=en-AU [accessed 2020-03-04]
16. Buller DB, Berwick M, Lantz K, Buller MK, Shane J, Kane I, et al. Smartphone mobile application delivering personalized, real-time sun protection advice: a randomized clinical trial. JAMA Dermatol 2015 May;151(5):497-504 [FREE Full text] [doi: 10.1001/jamadermatol.2014.3889] [Medline: 25629710]
17. Brinker TJ, Faria BL, de Faria OM, Klode J, Schadendorf D, Utikal JS, et al. Effect of a face-aging mobile app-based intervention on skin cancer protection behavior in secondary schools in Brazil: a cluster-randomized clinical trial. JAMA Dermatol 2020 May 6;156(7):737-745 [FREE Full text] [doi: 10.1001/jamadermatol.2020.0511] [Medline: 32374352]
18. Finch L, Janda M, Loescher LJ, Hacker E. Can skin cancer prevention be improved through mobile technology interventions? A systematic review. Prev Med 2016 Sep;90:121-132. [doi: 10.1016/j.ypmed.2016.06.037] [Medline: 27374946]
19. Hacker E, Horsham C, Vagenas D, Jones L, Lowe J, Janda M. A mobile technology intervention with ultraviolet radiation dosimeters and smartphone apps for skin cancer prevention in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2018 Nov 28;6(11):e199 [FREE Full text] [doi: 10.2196/mhealth.9854] [Medline: 30487115]
20. Hacker E, Horsham C, Ford H, Hartel G, Olsen CM, Pandeya N, et al. UV detection stickers can assist people to reapply sunscreen. Prev Med 2019 Jul;124:67-74 [FREE Full text] [doi: 10.1016/j.ypmed.2019.05.005] [Medline: 31077722]
21. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Mar;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
22. Arbon P. Mass-gathering medicine: a review of the evidence and future directions for research. Prehosp Disaster Med 2007;22(2):131-135. [doi: 10.1017/s1049023x00004507] [Medline: 17591185]
23. Safe Schoolies Gold Coast Response Information Guide. COVID-19 - Safer Schoolies (Queensland Government). 2019. URL: https://www.saferschoolies.qld.gov.au/resources/schoolies/information-guide.pdf [accessed 2020-05-06]
24. Lam T, Liang W, Chikritzhs T, Allsop S. Alcohol and other drug use at school leavers' celebrations. J Public Health (Oxf) 2014 Sep;36(3):408-416. [doi: 10.1093/pubmed/fdt087] [Medline: 23981443]
25. Barrett F, Usher K, Woods C, Harrison SL, Nikles J, Conway J. Sun protective behaviors at an outdoor entertainment event in Australia. Nurs Health Sci 2018 Mar;20(1):132-138. [doi: 10.1111/nhs.12396] [Medline: 29282864]

26. Mass Gathering Guideline. Gold Coast Hospital and Health Service. 2016. URL: https://www.goldcoast.health.qld.gov.au/sites/default/files/Mass%20Gathering%20Guidelines_0.pdf [accessed 2020-04-10]
27. 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]
28. Surfers Paradise Surf Cam & Surf Report. Coastalwatch: Surf Cams, Surf Reports and Surf Forecasts. URL: https://www.coastalwatch.com/surf-cams-surf-reports/qld/surfers-paradise [accessed 2020-04-20]
29. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, et al. ImageJ2: imageJ for the next generation of scientific image data. BMC Bioinformatics 2017 Nov 29;18(1):529 [FREE Full text] [doi: 10.1186/s12859-017-1934-z] [Medline: 29187165]
30. Glanz K, Yaroch AL, Dancel M, Saraiya M, Crane LA, Buller DB, et al. Measures of sun exposure and sun protection practices for behavioral and epidemiologic research. Arch Dermatol 2008 Mar;144(2):217-222. [doi: 10.1001/archdermatol.2007.46] [Medline: 18283179]
31. Heckman CJ, Handorf E, Darlow SD, Yaroch AL, Raivitch S. Refinement of measures to assess psychosocial constructs associated with skin cancer risk and protective behaviors of young adults. J Behav Med 2017 Aug;40(4):574-582 [FREE Full text] [doi: 10.1007/s10865-017-9825-3] [Medline: 28155000]
32. Pettigrew S, Jongenelis M, Strickland M, Minto C, Slevin T, Jalleh G, et al. Predictors of sun protection behaviours and sunburn among Australian adolescents. BMC Public Health 2016 Jul 13;16:565 [FREE Full text] [doi: 10.1186/s12889-016-3197-4] [Medline: 27411518]
33. Brinker TJ, Heckl M, Gatzka M, Heppt MV, Rodrigues HR, Schneider S, et al. A skin cancer prevention facial-aging mobile app for secondary schools in Brazil: appearance-focused interventional study. JMIR Mhealth Uhealth 2018 Mar 9;6(3):e60 [FREE Full text] [doi: 10.2196/mhealth.9794] [Medline: 29523502]
34. Perez D, Kite J, Dunlop SM, Cust AE, Goumas C, Cotter T, et al. Exposure to the 'dark side of tanning' skin cancer prevention mass media campaign and its association with tanning attitudes in New South Wales, Australia. Health Educ Res 2015 Apr;30(2):336-346. [doi: 10.1093/her/cyv002] [Medline: 25697580]
35. Tanning Teens Still Seek the Sun. Cancer Council: Cancer Information and Support. 2019. URL: https://www.cancer.org.au/news/media-releases/tanning-teens-still-seek-the-sun.html [accessed 2020-04-14]
36. Stapleton JL, Hillhouse J, Levonyan-Radloff K, Manne SL. Review of interventions to reduce ultraviolet tanning: need for treatments targeting excessive tanning, an emerging addictive behavior. Psychol Addict Behav 2017 Dec;31(8):962-978 [FREE Full text] [doi: 10.1037/adb0000289] [Medline: 28639816]
37. White KM, Zhao X, Sutton LC, Robinson NG, Hamilton K. Sun protection and young female beachgoers: a formative theory-based approach to identifying key sun safe beliefs. Health Promot J Austr 2019 Apr;30(2):263-266. [doi: 10.1002/hpja.189] [Medline: 30076670]
38. McWhirter JE, Hoffman-Goetz L. Systematic review of population-based studies on the impact of images on UV attitudes and behaviours. Health Promot Int 2015 Jun;30(2):397-410. [doi: 10.1093/heapro/dat031] [Medline: 23669159]
39. Half a Million Australian Teenagers Sunburnt on Summer Weekends. Cancer Council: Cancer Information and Support. 2019. URL: https://www.cancer.org.au/news/media-releases/half-a-million-australian-teenagers-sunburnt-on-summer-weekends.html [accessed 2020-03-04]
40. O'Hara M, Horsham C, Koh U, Janda M. Unintended sunburn after sunscreen application: an exploratory study of sun protection. Health Promot J Austr 2019 Oct 20 epub ahead of print. [doi: 10.1002/hpja.301] [Medline: 31631453]
41. Barrie L, Jones S, Lynch M, Coppa K. Confirming the existence and size of behavioural segments in the adolescent sun protection project: Results of a NSW schools survey. Australian and New Zealand Marketing Academy Conference, Melbourne, Victoria: Monash University 2009.
42. Robinson NG, White KM, Young RM, Anderson PJ, Hyde MK, Greenbank S, et al. Young people and sun safety: the role of attitudes, norms and control factors. Health Promot J Austr 2008 Apr;19(1):45-51. [doi: 10.1071/he08045] [Medline: 18481932]
43. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
44. Dobbinson S, Wakefield M, Hill D, Girgis A, Aitken JF, Beckmann K, et al. Prevalence and determinants of Australian adolescents' and adults' weekend sun protection and sunburn, summer 2003-2004. J Am Acad Dermatol 2008 Oct;59(4):602-614. [doi: 10.1016/j.jaad.2008.06.011] [Medline: 18691790]

Abbreviations

QUT: Queensland University of Technology
SEDs: standard erythemal doses
SPF: sun protection factor
SPH: sun protection habits

UVR: ultraviolet radiation

Edited by G Eysenbach; submitted 10.06.20; peer-reviewed by K Dixon, D Stratton, M Jongenelis; comments to author 09.07.20; revised version received 21.07.20; accepted 23.07.20; published 16.09.20.

Please cite as:
Horsham C, Antrobus J, Olsen CM, Ford H, Abernethy D, Hacker E
Testing Wearable UV Sensors to Improve Sun Protection in Young Adults at an Outdoor Festival: Field Study
JMIR Mhealth Uhealth 2020;8(9):e21243
URL: http://mhealth.jmir.org/2020/9/e21243/
doi:10.2196/21243
PMID:32936083

Original Paper

Breast Cancer Survivors' Perspectives on Motivational and Personalization Strategies in Mobile App–Based Physical Activity Coaching Interventions: Qualitative Study

Francisco Monteiro-Guerra[1,2], MSc; Gabriel Ruiz Signorelli[2,3], MSc; Octavio Rivera-Romero[4], PhD; Enrique Dorronzoro-Zubiete[4], PhD; Brian Caulfield[2], PhD

[1]Salumedia Tecnologías, Seville, Spain

[2]The Insight Centre for Data Analytics, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland

[3]Oncoavanze, Seville, Spain

[4]Department of Electronic Technology, Universidad de Sevilla, Seville, Spain

Corresponding Author:
Francisco Monteiro-Guerra, MSc
Salumedia Tecnologías
Avda. República Argentina
nº 24 Edificio Torre de los Remedios, 5ª planta - módulo A
Seville, 41011
Spain
Phone: 34 672787858
Email: francisco.guerra@ucdconnect.ie

Abstract

Background: Despite growing evidence supporting the vital benefits of physical activity (PA) for breast cancer survivors, the majority do not meet the recommended levels of activity. Mobile app–based PA coaching interventions might be a feasible strategy to facilitate adherence of breast cancer survivors to the PA guidelines. To engage these individuals, PA apps need to be specifically designed based on their needs and preferences and to provide targeted support and motivation. However, more information is needed to understand how these technologies can provide individual and relevant experiences that have the ability to increase PA adherence and retain the individual's interest in the long term.

Objective: The aim of this study is to explore insights from breast cancer survivors on motivational and personalization strategies to be used in PA coaching apps and interventions.

Methods: A qualitative study was conducted, using individual semistructured interviews, with 14 breast cancer survivors. The moderator asked open-ended questions and made use of a slideshow presentation to elicit the participants' perspectives on potential mobile app–based intervention features. Transcribed interviews were evaluated by 3 reviewers using thematic content analysis.

Results: Participants (mean age 53.3, SD 8.7 years) were White women. In total, 57% (8/14) of the participants did not adhere to the PA guidelines. In general, participants had access to and were interested in using technology. The identified themes included (1) barriers to PA, (2) psychological mediators of PA motivation, (3) needs and suggestions for reinforcing motivation support, (4) personalization aspects of the PA coaching experience, and (5) technology trustworthiness. Motivational determinants included perceived control, confidence and perceived growth, and connectedness. Participants were interested in having a straightforward app for monitoring and goal setting, which would include a prescribed activity program and schedule, and positive communication. Opinions varied in terms of social and game-like system possibilities. In addition, they expressed a desire for a highly personalized coaching experience based on as much information collected from them as possible (eg, disease stage, physical limitations, preferences) to provide individualized progress information, dynamic adjustment of the training plan, and context-aware activity suggestions (eg, based on weather and location). Participants also wanted the app to be validated or backed by professionals and were willing to share their data in exchange for a more personalized experience.

Conclusions: This work suggests the need to develop simple, guiding, encouraging, trustworthy, and personalized PA coaching apps. The findings are in line with behavioral and personalization theories and methods that can be used to inform intervention design decisions. This paper opens new possibilities for the design of personalized and motivating PA coaching app experiences for breast cancer survivors, which might ultimately facilitate the sustained adherence of these individuals to the recommended levels of activity.

(JMIR Mhealth Uhealth 2020;8(9):e18867) doi:10.2196/18867

KEYWORDS

mHealth; mobile app; mobile phone; coaching; physical activity; breast cancer

Introduction

Background

Breast cancer affects a large and growing population of women globally and is the second leading cause of death among women [1,2]. Fortunately, owing to advancements in screening and treatment, survival rates are on a steady rise [3]. In their journey during and after treatment, breast cancer survivors often need personalized support and encouragement to take a proactive approach in the self-management of their health and quality of life [4]. Participating in physical activity (PA), such as walking or jogging, is recommended as part of that process providing vital benefits to survivors, which include prevention of cancer recurrence, decrease in side effects from treatment, and improvements in fitness, body size, and quality of life [4-6]. However, most breast cancer survivors do not meet the PA guidelines and recommendations, with reported estimations of <10% in some studies based on self-reported PA data [7,8].

The growing demand for cost-effective alternatives to help people reach their personal activity targets, paired with the advancements in mobile monitoring technologies, has increased interest in the research and development of mobile PA coaching apps and interventions [9-11]. These are aimed at motivating people to change their activity behavior by means of coaching elements, such as measuring the user's daily PA and providing feedback, guidance, and incentives, to generate awareness on current behavior and ultimately stimulate users to achieve and maintain an active lifestyle [12]. At present, these systems have the potential to provide engaging real time coaching experiences for users, and as suggested in related studies, these systems may be a feasible strategy for motivating cancer survivors to adhere to the recommended levels of PA [13,14].

Although there has been an increase in the popularity of PA apps, the vast majority is targeted at the healthy active population [15]. These apps are generally not suitable for individuals who may struggle with disease-specific barriers [16] and may have particular coaching needs for PA, such as breast cancer survivors [14]. To increase the likelihood of mobile PA coaching interventions being accepted and benefitting breast cancer survivors, it is essential to understand their contextual characteristics, needs, and preferences as end users of the technology to inform system design [17]. A few studies have looked into the specific requirements of breast cancer survivors for technology-based PA coaching interventions [14,18,19]. However, there is more to understand about what can hold their interest and motivate them to increase their PA levels in the long term. Previous work highlights the importance of exploring behavioral theories of motivation and personalization aspects in the design of these systems [20].

Using methods from behavioral theory in tools that aim to increase PA is believed to increase the success of these technology-supported interventions [21]. To achieve this, the technology should include evidence-based techniques drawn from behavior change theories that directly address the special barriers and motivators for PA adherence among the target users. Previous studies suggest that theories such as the social cognitive theory (SCT), transtheoretical model (TTM), and self-determination theory (SDT) should be used in the design of PA interventions for breast cancer survivors [22,23]; however, there is more to learn in this direction and to have a better understanding of the practical implications of technology implementation [23,24]. Furthermore, it is generally believed that tailoring or personalizing *any number of methods for creating communications individualized for their receivers* is associated with an increase in the effectiveness of behavior change systems [25,26]. Personalization can contribute to captivating and holding the user's interest, which is of utmost importance in systems that generally struggle with user abandonment [27-29]. Recent work in this area deals specifically with real time personalization of PA coaching apps and suggests a number of strategies (eg, user targeting, goal setting, adaptation) that can be used and combined for that purpose [12,20]. In line with this, previous studies with breast cancer survivors seem to suggest a preference for an app experience that is highly tailored to each user [14,24,30]. However, there is still very little evidence on which factors can be taken into account for the creation of individualized experiences for these individuals.

Objectives

To our knowledge, no study has thoroughly explored the perspectives of breast cancer survivors toward personalization and motivational aspects in PA coaching apps. Therefore, we conducted a qualitative research study with the goal of obtaining practical insights from breast cancer survivors on how a PA app may be accepted and provide personally relevant and engaging experiences to these individuals. These insights were applied to established constructs from theory on behavior change, SCT, and SDT and on personalized coaching to inform future app design.

Methods

Study Design

This study adopts a qualitative approach comprising semistructured interviews. The methods used were based, in part, on the qualitative work of Robertson et al [13] on mobile health (mHealth) PA intervention preferences in cancer survivors.

Individual face-to-face semistructured interview sessions were conducted to gather information on the patients' experiences and perspectives about PA coaching apps, including their thoughts and experiences (eg, barriers and facilitators or motivators) about PA and the use of PA apps, and their insights on various coaching strategies and characteristics.

Theoretical Basis

To understand the determinants of PA among breast cancer survivors, we considered 2 theoretical approaches that address this issue of motivation: the SCT and the SDT. The SDT by Ryan and Deci [31] suggests that self-motivation evolves from the degree to which a person's innate psychological or basic needs are met within their social context. The 3 psychological needs identified in the theory are competence, a person's ability to interact effectively within their environment; autonomy, a person's perceived control of their choices; and relatedness, sense of connectedness to others in the immediate environment [32]. The theory proposes that satisfying these needs may promote and facilitate motivation by developing more intrinsic motivations—performing tasks for inherent enjoyment, interest, and pleasure of accomplishing them—consequently leading to task persistence [31]. The SCT, as proposed by Bandura et al [33,34], establishes that a person with given beliefs, information, and needs functioning in given social and physical environments will engage in a behavior that will have a consequent outcome. The 2 primary determinants of behavior in SCT are expected outcomes: expectations about the outcomes resulting from engaging in behavior and self-efficacy, or efficacy, expectations (beliefs) about one's ability to engage in or execute the behavior. Self-efficacy is widely recognized as one of the strongest determinants of PA participation [35]. Another core component of SCT is the sociostructural factor, which includes impediments and facilitators to performing the desired behavior.

Another aspect that has been associated with an increase in engagement and effectiveness of behavior change systems is personalization, which may help increase the intended effects of the communication provided by these technologies. In particular, the model of real time personalization in PA coaching systems [12] defines 7 different personalization concepts that can be explored in these systems to provide unique and dynamic experiences to each user. They are feedback, presenting individuals with descriptive, comparative, and/or evaluative information about themselves; goal setting, presenting the user with short-term and long-term goals that can create a feeling of progress; user targeting, conveying, explicitly or implicitly, that the communication is designed specifically for the user; interhuman interaction, support for any form of interaction with other real human beings; adaptation, direct communication to individuals' status on key theoretical determinants of the behavior of interest; context awareness, providing relevant information and/or services to the user based on the user's context (not including user characteristics) and needs or goals; and self-learning, the ability to update the internal model of the user over time by recording and learning from the various user interactions with the app.

Finally, gamification, *the use of game design elements in nongame contexts* [36], was also considered in the analysis of this study. Gamification has been broadly used in health and fitness apps, with studies suggesting its potential to create fun and engaging experiences for the users [37,38]. Gamification methods include points, rewards, competition, avatars, and themes.

Recruitment

The inclusion criteria for research participants were oncology patients with a history of breast cancer (stages I-III) that finished primary curative treatment (ie, surgery, radiotherapy, and chemotherapy), aged>18 years, own and use a mobile phone or smartphone, and have the ability to read and speak Spanish, with no known impairments or comorbidities, and no restrictions on PA.

The participants were recruited from the Oncology Clinic of the Oncoavanze group (in Seville, Spain) by placing a phone call to the eligible individuals identified in the Oncoavanze patient database. Recruitment was conducted until saturation of results was reached, which was considered when there was no new information (themes) arising from the qualitative data.

This study was approved by the Biomedical Research and Ethics Committee of the Junta de Andalucia in Spain. Subjects' agreement for participation was obtained through an informed consent process.

All individuals who decided to participate had attended the interview sessions, and none of them dropped out of the study.

Data Collection

The study sessions took place in a private room at the Oncoavanze headquarters, where only the session facilitator and the participant were present. Data collection ran from July 12 to 27, 2018, and was conducted by the main investigator (FG), a biomedical engineer and PhD candidate trained in qualitative research methods. The sessions lasted 35-60 min.

Standardized Questionnaires and Demographics

At the beginning of the sessions, initial assessments were made to describe the participants' characteristics. Structured questionnaires and scales were administered to collect data on demographics (questionnaire), satisfaction with life scale (SWLS; Spanish version of the SWLS [39]), technology use and interest (questionnaire based on a study by Robertson et al [13]), and PA (questionnaire: Spanish version of the International Physical Activity Questionnaire-Short Form, IPAQ-SF [40]). The IPAQ-SF results were also used to extrapolate the adherence of participants to the PA guidelines [41], a step done by GS.

Semistructured Interviews

During the interviews, FG provided trigger questions to the participants based on a semistructured interview guide and took field notes. The first part of the interview was a discussion in which the moderator asked open-ended questions, followed by conversational probes as appropriate. This part included questions about participants' experiences with managing their health and PA, including barriers and facilitators, previous use of mobile apps, and perspectives toward a tailored PA app and related requirements (Multimedia Appendix 1). During the second part, after a general introduction about PA apps and personalization, the moderator used a slideshow presentation of possible app features (Multimedia Appendix 2) and asked participants to share their thoughts and opinions on the featured content. In an attempt to avoid obtaining insights biased by the example features shown, before showing each of them, the

moderator asked participants if they could contextualize the strategy being discussed with a real life example not necessarily related to PA or apps (eg, can you think of an example in your life where defining goals was important or useful to you?).

Two researchers, FG and OR, designed the interview guide following an iterative process (Multimedia Appendix 1). The questions and slideshow presentation were derived from the work of Robertson et al [13] and chosen to combine concepts of personalization in PA coaching apps [12,20] and behavioral theory [31,34] (eg, monitoring, goal setting, social connectedness, and context awareness) as well as practical questions (eg, participant perspectives and preferences). Given that the literature suggests that breast cancer survivors' main and preferred form of exercise is walking, including activities of daily living [18,24], and to have a more comprehensive research in that direction, the features presented targeted specifically such types of aerobic activities. In addition, the slideshow presentation and explanations were designed assuming an audience with low technical knowledge.

Interviews were conducted in Spanish. All sessions were recorded with a digital audio recorder and transcribed verbatim. All transcribed data were deidentified and securely stored on password-protected computers.

Participants' quotes considered in the analysis were later translated to English for writing the Results section. It was not a literal translation owing to the amount of unique linguistic expressions used by the participants, which are characteristic of their region. Translated transcripts were validated linguistically and culturally through back-translation techniques and evaluated by bilingual translators.

Data Analysis

Data analysis was conducted by 3 independent reviewers (FG, OR, and ED) on the original Spanish transcripts, and consisted of both deductive and inductive approaches. FG is a PhD student trained in qualitative research methods, and both OR and ED are senior researchers with experience in qualitative research.

A preliminary coding frame was defined by 2 researchers (FG and OR) based on the main topics of interest assessed using the interview guide: barriers and facilitators on PA, insights on previous app use, perspectives on personalization, and opinions on potential app functionalities and characteristics. This initial deductive phase allowed the coders to become familiar with the content and filter out irrelevant information.

After this initial step, the 2 coders (OR and ED) used an approach drawn from inductive methods, based on a thematic content analysis [42], and theoretically informed deductive methods. In the latter method, SCT [33,35,43] and SDT [31,44] constructs, personalized coaching strategies [12,20,25], and gamification mechanics [37] helped to identify internal and external motivational influences on participants' PA adherence and PA app use. The analysis involved a comparison of the aforementioned theories with the raw data. This comparison enabled the researchers to code theoretically relevant constructs (eg, competence, relatedness) that then helped to identify themes associated with motivational determinants of PA adherence (eg, confidence and perceived growth, social connectedness), intervention characteristics (eg, activity monitoring and goal setting, interacting with other users), and how they can support motivation by influencing the motivational determinants.

At the end of each iteration, the themes and subthemes were reviewed and refined. Recurring themes and subthemes were consolidated and coded through an iterative process.

Preliminary findings were then reviewed by a third coder (FG), who reviewed it against the interview transcripts and field notes and proposed a final readjustment to the results. Any discrepancies during the coding process were resolved by consensus. Illustrative quotes for each subtheme were selected from the data set.

Microsoft Excel version 16 was used to organize content within thematic categories.

Results

Participant Characteristics

Semistructured interviews were conducted with 14 patients with breast cancer. The age of the participants ranged from 43 to 69 years, with a mean of 52.8 (SD 8.8) years and a median of 50 (IQR 47.3-56.8) years. The number of years since diagnosis ranged from 1 to 11.5, with a mean of 4.25 (SD 2.8) years and a median of 3.9 (IQR 2.5-5.4) years. Participants were well educated and most were employed (11/14, 79%). The IPAQ-SF scores indicated that most participants had a moderate or high level of PA (12/14, 86%). When analyzing the IPAQ-SF answers in terms of the activity type and intensity, 8 participants reported only doing low-intensity activities (eg, light walks) and therefore did not adhere to the PA guidelines. In addition, 7 of the 14 participants reported spending 5 hours a day of sedentary behavior. The mean SWLS score was 17.6 (SD 4.2), indicating a neutral-to-good satisfaction with life. Most participants reported being very interested in technology, reported having ready access to technological devices, and have shown high usage of a variety of technology functionalities (Multimedia Appendix 3). However, some reported neutral self-reported skill with technology (5/14, 36%). More details are provided in Table 1.

Table 1. Participant characteristics (N=14).

Characteristics	Values, n (%)
Marital status	
Single	4 (29)
Married	10 (71)
Divorced	0 (0)
Education	
Basic school	1 (7)
High school	2 (14)
Higher education	2 (14)
University or college	9 (64)
Current employment status	
Not working	3 (21)
Employed	11 (79)
Receiving pharmacological treatment	10 (71)
Indication for PA[a]	11 (79)
IPAQ-SF[b] level	
High	1 (7)
Moderate	11 (79)
Low	2 (14)
Adheres to PA guidelines (>150 min per week=moderate activity or >75 min per week=vigorous activity)[c]	6 (43)
Interest in technology	
Agree or strongly agree	12 (86)
Neutral	2 (14)
Disagree or strongly disagree	0 (0)
Self-reported skill with technology	
Agree or strongly agree	9 (64)
Neutral	5 (36)
Disagree or strongly disagree	0 (0)
"I like to experiment with new technology"	
Agree or strongly agree	7 (50)
Neutral	5 (36)
Disagree or strongly disagree	2 (14)

[a]PA: physical activity.

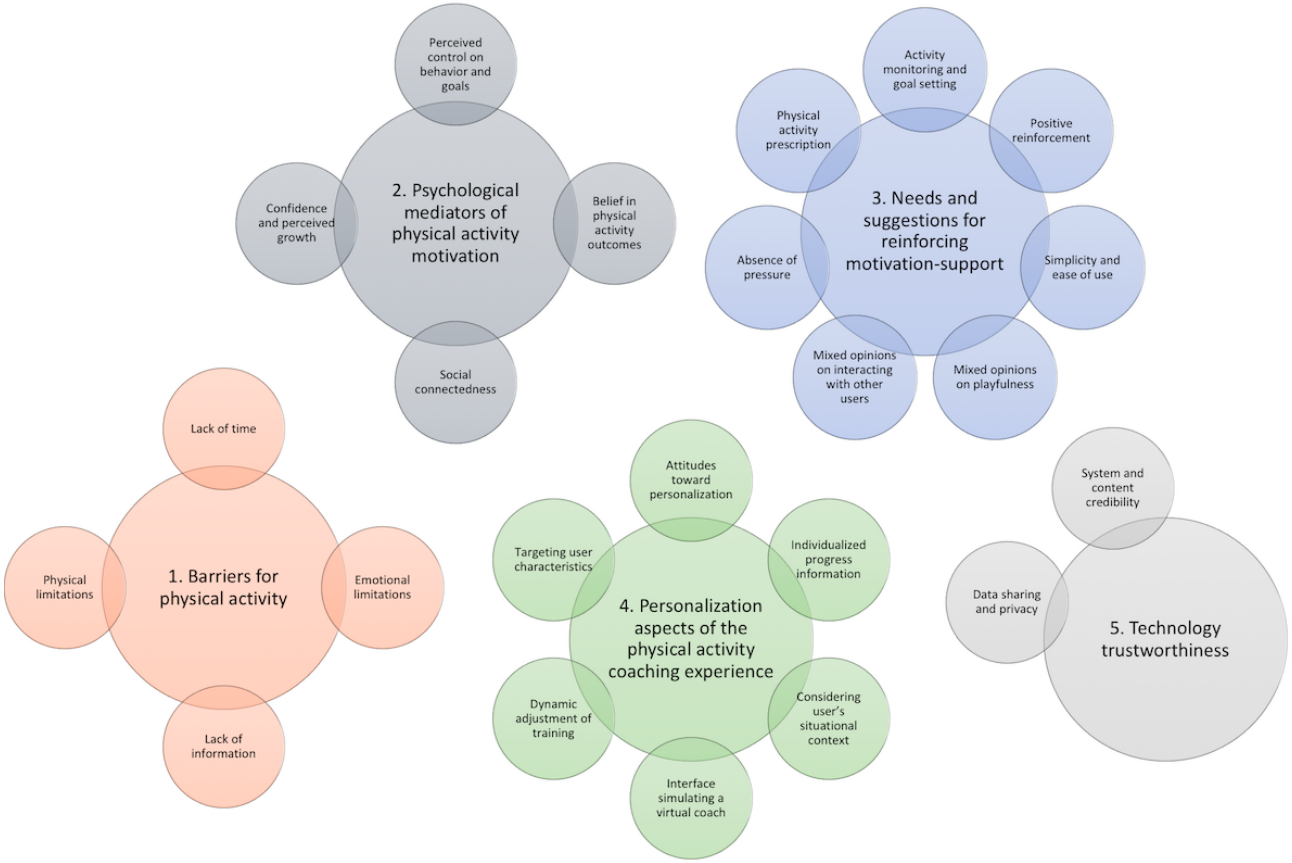
[b]IPAQ-SF: International Physical Activity Questionnaire-Short Form.

[c]On the basis of the IPAQ-SF answers.

Themes

The following themes were identified from the interview data. These include aspects related to PA adherence and design considerations for PA apps in breast cancer survivorship. The main themes identified were as follows: (1) barriers to PA, (2) psychological mediators of PA motivation, (3) needs and suggestions for reinforcing motivation support, (4) personalization aspects of the PA coaching experience, and (5) technology trustworthiness (Figure 1). Subthemes are introduced below within each of the main themes' sections. A table with supporting quotation is provided in Multimedia Appendix 4.

Figure 1. Schematic representation of themes and subthemes.



Barriers for PA

Although participants seemed generally aware of the importance of PA, they often expressed the challenges and difficulties that prevent them from being more active. These included lack of time, physical limitations, emotional challenges, and lack of information. From the SCT perspective, these can be defined as sociostructural factors that can inhibit the performance of the behavior.

Lack of Time

Time management challenges were often raised across most participants. The 2 main reasons associated with these challenges were work schedule and being in charge of relatives:

> *My daughters are still ten years old, and during the morning I work, I take them to school and I pick them up from school. So, when I don't have to take them to school, I walk [to work]. […]. At work, I am sitting most of the morning. And, in the afternoon, I combine their extracurricular activities [in own schedule], so I'm not very constant in maintaining physical activity.* [P03, aged 48 years, moderate PA level, high skill with technology]

> *We live so fast that we don't know how to manage time, and [then] we say we don't have time to do it. […] The barrier is time management because I'm in an important professional moment […] and on top of that I'm also a mum, so I choose [to leave aside] sports… I leave it [sports] as a last priority […].* [P10, aged 39 years, moderate PA level, neutral skill with technology]

Physical Limitations

Many breast cancer survivors report experiencing side effects from the disease and treatment, particularly physical problems, which can limit the extent to which they participate in PA. Among the physical burdens reported by the participants were lymphedema, weight gain, changes in body image, muscular and joint pain, fatigue, and neuropathy:

> *Because you know that due to the "remains" of the neuropathy, fatigue, tiredness, and muscular pains, either you move or everything [all the physical burdens] will be much worse... it will hurt more or you will be more tired.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

> *[I felt] many barriers associated with the lymphedema and weight [gained], when doing activity [exercise] with gym equipment.* [P05, aged 43 years, moderate PA level, high skill with technology]

> *[...] between what they put in you [referring to medication] and the volume of weight that you gain [...] you get out of breath.* [P08, aged 47 years, moderate PA level, high skill with technology]

Emotional Challenges

Some participants suggested that their emotions were also affected after treatment and that it could negatively influence their motivation to participate in PA. Such emotional instability was reflected in feeling more stressed and depressed. They often

suggested how common it was to have bad thoughts after the cancer:

> *We are people who stress easily and manage stress even worse now than before the disease. Also because of the medication we receive... it gives some emotional instability…* [P02, aged 50 years, moderate PA level, neutral skill with technology]

> *For a while now I have stopped doing it [physical activity] because I was a little depressed and stressed.* [P04, aged 48 years, moderate PA level, high skill with technology]

> *People start [saying] "Oh, I'm feeling bad, I'm feeling bad, I'm feeling bad...", and in that way, you don't recover.* [P04, aged 48 years, moderate PA level, high skill with technology]

Lack of Information

Although most participants were advised to be more active by their health care professionals, they reported having unmet information needs on the type, amount, and duration of PA. It was also suggested that some doctors were outdated and recommended rest instead of PA:

> *[...] the order of exercises or of stretching [...] or, maybe, if something hurts, what exercises can be better for you, or what type of things you shouldn't do.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

> *What makes me angry is that there are many doctors that are still very outdated and the first thing they tell you is to rest… that's the easiest… and then people "rust".* [P04, aged 48 years, moderate PA level, high skill with technology]

> *[...] you often find yourself a bit disoriented [...]. [Talking about mastectomy] I found out much later, once I had already finished the treatment, about lymphedema… I did not know that I had to do some exercises.* [P07, aged 56 years, moderate PA level, high skill with technology]

Psychological Mediators of PA Motivation

This theme presents the psychological mediators of behavior change associated with breast cancer survivors' adherence to PA. These were analyzed through the lenses of the SDT and SCT approaches and included perceived control on behavior and goals, confidence and perceived growth, belief in PA outcomes, and social connectedness.

Perceived Control of Behavior and Goals

Perceived control is related to the concept of autonomy, as defined in the SDT, which is associated with the feeling of being the origin of one's own behaviors and experiencing volition in action. Participants reported a willingness to play an active role in managing their health as it could positively influence their own and their loved ones' lives. This can be associated with a sense of ownership and value alignment, which influence perceived autonomy:

> *It was me [the source of motivation]. Me, because I had to save my children… I was pregnant.* [P04, aged 48 years, moderate PA level, high skill with technology]

> *I have been given an opportunity of being here, in life, again... and I have to make the most out of it, doing good for others and being happy.* [P11, aged 69 years, moderate PA level, high skill with technology]

In addition, there seemed to be a need to self-regulate and feel in control of their PA experience. Participants highlighted several aspects related to acting in alignment with their own goals and choosing a personal PA strategy, which are characteristics of autonomy support:

> *The objective is to always have some group or activity to sign up to, always… be it pilates or aquagym. [...] not leaving it to when I feel like doing it, but to present myself with a concrete activity. The specific objective is general maintenance, controlling… I do it for my health [...]* [P01, aged 57 years, low PA level, neutral skill with technology]

> *[...] Now I'm more interested in the subject of calories, due to the pill that I'm taking, anti-hormone, and all that. I want to have more control over it. [...] Setting goals and objectives is always important. [...]* [P13, aged 47 years, moderate PA level, high skill with technology]

> *I think that in the end it's you who sets your own targets and if you feel good with what you are doing then that's all you need.* [P07, aged 56 years, moderate PA level, high skill with technology]

Confidence and Perceived Growth

A person's confidence in one's own capacity to perform a behavior or a task is defined, according to SCT, as self-efficacy. It is often associated with the concept of competence of SDT, which is about the need for perceived capability and growth in performing the desired behavior. In the context of breast cancer survivors, the belief in their capacity to perform more PA seemed to be limited by their physical and psychological barriers and fears. Participants referred to their difficulties in starting, or restarting, to perform PA after treatment and how they needed to start slow. There was this general perception that they had lower physical capabilities than before treatment:

> *I went back to training and you want to start almost where you left off, and physically you are not in the same conditions, then it costs you a little, it's like starting from scratch.* [P05, aged 43 years, moderate PA level, high skill with technology]

> *In the beginning, when I finished the treatments, it was very difficult to do physical activity. […] I had more fear than actual [self-] confidence […].* [P02, aged 50 years, moderate PA level, neutral skill with technology]

Participants talked about the importance of feeling improvements, for example being progressively challenged, being successful in achieving defined goals, and feeling

improvements on their biological processes (eg, weight loss and heart rate). This is related to the concepts of growth and mastery of experience, from SCT and SDT, which involve exercising one's capacities, acquiring new skills, or receiving constructive feedback. These, over time, are believed to increase confidence, autonomy, and satisfaction:

> *So, you realize how you are improving, because, of course throughout the time that you are burning calories, you are losing weight and you are improving your performance.* [P04, aged 48 years, moderate PA level, high skill with technology]

> *[That] these objectives would vary if you have been accomplishing part of them… that they would be each time more complex and allowing you to overcome yourself. So, in this way, […] you start feeling better because every time you can achieve more objectives, doing more stuff or goals.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

Some participants also suggested that their source of confidence was associated with their success in overcoming their disease or relapses. In this line, having positive and constructive thoughts was an important part of their process of acceptance and self-motivation after cancer. This is related to the idea of positive self-talk in SCT, which is believed to increase self-efficacy (confidence):

> *I think it motivated me psychologically that I could say "I did it [recovered] and I was capable of doing it", and it was extremely complicated… [it was] a pretty hard road [process of recovery].* [P05, aged 43 years, moderate PA level, high skill with technology]

Belief in PA Outcomes

Interviewees seemed to be aware of the importance of integrating PA in their lives. This can be associated with the concept of outcome expectations from SCT. Participants' awareness of the physical and psychological benefits of PA in breast cancer survivorship was highlighted as a source of motivation to be more active:

> *The fact that you can move more, helps you relax your mind. The fact that you can find yourself more agile, helps you to feel better about yourself… [...] I believe sport is fundamental [...], it's about constant improvement… which, for people who come out of cancer [...], and that have been in a capsule of medicines, pain and mental focus [...], when that capsule opens you are so broken that any reasonable target is seven or ten steps that you climb… sport helps a lot in that sense.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

> *It motivates me that I feel better [by being active]… I feel much better physically and psychologically. More lively, as if with more strength, energy… yes, that's what motivates me.* [P05, aged 43 years, moderate PA level, high skill with technology]

Social Connectedness

Social connectedness, or relatedness, is one of the psychological needs of SDT. Aspects related to social support by health care professionals, peers, and close ones were raised during the interviews. Participants had varied opinions on relatedness depending on the source of support. These aspects are also related to the component of SCT on sociostructural factors that can affect behavior.

Health Care Professionals

Opinions on interactions with health care professionals who took part in their treatment were generally very positive. Often, participants stated that they were crucial in helping them become motivated to cope with the disease and to adhere to PA, and particularly highlighted the importance of the psycho-oncologists:

> *The psychologist is very important because you talk to her [...], she cheers you up and then you feel like coming back and do stuff [talking about PA].* [P05, aged 43 years, moderate PA level, high skill with technology]

> *[After treatment] I had more fear than actual [self-] confidence. I was lucky to go to a psycho-oncologist [...] and she encouraged me not to stop doing physical activity… I also met [name of an exercise trainer] [...] which was a very important help.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

Peers

Opinions on peer support and interaction were not straightforward, with a mixture of positive and negative thoughts by participants. On the positive side, because breast cancer survivors share this common life experience, they can understand each other's situations, feelings, and fears, which enhances the feeling of relatedness and facilitates positive interactions between them:

> *If you get together with someone that has a similar experience, and you talk and share your feelings, that's a support.* [P01, aged 57 years, low PA level, neutral skill with technology]

> *Especially when you are with people who are going through the same as you, then you vent a lot [...] sometimes you don't have to talk, just hanging out, laughing, and disconnecting from problems... it helps a lot.* [P05, aged 43 years, moderate PA level, high skill with technology]

Positive opinions regarding performing PA in groups were also expressed by participants. In particular, it was suggested that it could help in socializing and avoiding getting bored. In addition, the fact that they could see others like themselves performing PA seemed to be motivating:

> *Doing PA in a group is different… but if you are doing alone, you do not progress, and sometimes you give up because you get bored. When you do not progress, it is interesting if someone motivates you*

> *in some way.* [P05, aged 43 years, moderate PA level, high skill with technology]
>
> *I think that another very funny thing about activities is to be able to socialize with other people.* [P07, aged 56 years, moderate PA level, high skill with technology]
>
> *Seeing how [the trainer] trained other women and see how she made them swim and do exercises… to me, that was important, truly.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

Negative feelings toward peer interactions were based on the potential negative impact that it could have on their own emotions. Some participants reported an almost unavoidable tendency to speak about their cancer experiences, often negative and sad, in peer groups, which could prevent them from potentially benefiting from being part of these groups. In some cases, this was accentuated by their health professionals who recommended them to avoid such interactions at certain stages of the treatment. In addition, group heterogeneity in terms of PA levels was also suggested by participants as a barrier to adherence:

> *[People of the group] were talking about the disease, the types of intervention, etc.: "This doctor is not good" and then it turns out they were talking about your doctor; "This type of intervention no" and it turns out it was yours [...]. [Name of the oncologist] once told me "don't interact with anyone... interact with people who don't have cancer.*" [P03, aged 48 years, moderate PA level, high skill with technology]
>
> *But, in that case [talking about a Nordic Walking group], it was my own [physical] limit [as the barrier]... because I was in one homogeneous group [with higher PA level] and I was asphyxiated [couldn't handle the activity level].* [P03, aged 48 years, moderate PA level, high skill with technology]

Family and Friends

Related ones had a particularly relevant role in providing support for participants to better manage their health, being identified as an important source of motivation. Some participants reported a very positive influence on their families' involvement in their PA experiences:

> *I try to share these goals with the people around me so that they know [my goals] and they can help me, as a support network.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

Needs and Suggestions for Reinforcing Motivation Support

This theme identifies a number of strategies and characteristics that, from the participant's perspective, could address their motivational needs for PA: activity monitoring and goal setting, PA prescription, positive reinforcement, absence of pressure, simplicity and ease of use, mixed opinions on playfulness, and mixed opinions on interacting with other users.

Activity Monitoring and Goal Setting

Receiving information on the activity performed, together with goal setting, may help monitor the desired behavior and take actions to regulate it. Both monitoring and regulating are the main components of self-control, which is a construct used in SCT-based interventions. The participants were very interested in being able to monitor the different aspects related to their activities (including steps, duration, distance, and speed). In addition, some were particularly interested in knowing the calories burnt during exercise:

> *All the information [regarding own PA] seems important to me… the more information you have on an exercise that you are doing, the better. Everything seems important because sometimes you are guided [by this information].* [P05, aged 43 years, moderate PA level, high skill with technology]
>
> *I find it super motivating… [for example] if yesterday I ran an hour and I walked so many steps, [that the app shows] these many kilometers or these many [calories] burnt. I find it super interesting.* [P10, aged 39 years, moderate PA level, neutral skill with technology]

Interviewees often highlighted the importance of clearly defining goals to help them adhere to PA. Goal setting is a behavior change strategy that fits and can be inspired in SCT and SDT. Suggested goals were mostly quantitative and short term, for example, a number of steps, distance, duration, or speed. Participants also mentioned that these should be slightly more challenging each time:

> *It is very common in my life to set objectives in the short-term, particularly now, and some more on the long-term, but never much on the long-term... not too much… a period that makes sense to me. Yes, I do this [setting goals] a lot.* [P02, aged 50 years, moderate PA level, neutral skill with technology]
>
> *What I've proposed during this process [of trying to be more active] was "I have to do this many daily steps, and each day I will be better... each day I will do a little more".* [P06, aged 50 years, moderate PA level, neutral skill with technology]
>
> *To me, that I'm a very organized person, it sounds very good because it's a way of controlling, you control what you are doing [in terms of activity].* [P13, aged 47 years, moderate PA level, high skill with technology]

PA Prescription

Participants suggested having a prescribing tool that would help them plan their PA (key component of goal setting), with details on what they have to do and when to do it. It was also ideated as an activity program that would help them recover their physical fitness and routine after treatment. This seemed to be a way of addressing their barriers of lack of time and lack of detailed PA information:

> *What I had told you before, having a calendar or creating a schedule [in a physical activity app] would*

> *be great.* [P01, aged 57 years, low PA level, neutral skill with technology]
>
> *[…] there is a lot of people who don't know how to plan. It seems great to have a plan of what you have to do and when you have to do it [physical activity], from Monday to Sunday… It seems perfect.* [P10, aged 39 years, moderate PA level, neutral skill with technology]
>
> *[…] you finish treatment and you have abandoned a little exercising [...] and physically you are not the same. Better [to have] that, a program that tells you how to resume everything, how to come back to training.* [P05, aged 43 years, moderate PA level, high skill with technology]

Some also suggested including recommendations for other types of exercises, not only physical but also for relaxation, meditation, or nutrition:

> *[...] in those moments that you are not feeling well, maybe you have to do something of relaxation or meditation for those alternative days [...].* [P06, aged 50 years, moderate PA level, neutral skill with technology]

Positive Reinforcement

Reinforcement is a construct that is widely used in SCT-based interventions. Reinforcements are defined as incentives and rewards to encourage behavior change. Participants reported the importance of receiving such positive reinforcements to acknowledge their efforts:

> *[...] with positive reinforcement, for example, "look how well you did" and "you have completed your daily objective" or "you have little left to achieve your weekly goal" [...]. So, for me [it would be enough] a simple recognition of "you did well", [or] a funny and convincing "you're on your way" [...] It is great [to have this].* [P02, aged 50 years, moderate PA level, neutral skill with technology]
>
> *When you reach your objective that, in some form of message, acknowledges it and incites you to continue, and to propose new challenges.* [P06, aged 50 years, moderate PA level, neutral skill with technology]

Absence of Pressure

Another relevant characteristic to some participants was the absence of pressure, as they did not want to feel pressure toward performing PA. This was particularly evident when talking about app communication. Participants reported not wanting to get negative reactions from the app when they did not reach their goals. This is associated with the characteristics, positive feedback from competence support, and absence of pressure from autonomy support:

> *I wouldn't like that it would react [...] like "you have failed". So, what I mean is that the [virtual] coach [when the user does not comply with an objective] would never react with a negative message. [...] [Instead] it should be "cheer up, I will wait for you tomorrow at 9".* [P10, aged 39 years, moderate PA level, neutral skill with technology]
>
> *[…] if after the goals are not reached, there is no need to get frustrated.* [P13, aged 47 years, moderate PA level, high skill with technology]

In addition, participants suggested wanting the *just enough* reminders or notifications from the app:

> *Well, I think it's quite interesting, if it doesn't bombard you too much. Imagine that you don't feel like it, and there's this annoying thing telling you [to do it].* [P01, aged 57 years, low PA level, neutral skill with technology]
>
> *That would be very good. [...] A reminder to get a move on the physical matter, [...] to push you to get up. [...] But every day disturbing, no.* [P11, aged 69 years, moderate PA level, high skill with technology]

Simplicity and Ease of Use

Participants wanted a straightforward and easy-to-use app. Some reported having difficulty using complex apps and highlighted the need for a simple, visual, and intuitive interface that was easy to learn and very well explained. Making the system easy to operate can support a sense of competence in users:

> *Many times I uninstall many apps because they are complex and it takes me time that I don't want to invest, to learn how to use them. I do not like to spend time to learn it. I prefer something more simple and that later can turn into something more complicated [...].* [P02, aged 50 years, moderate PA level, neutral skill with technology]
>
> *[...] simple, that it would be well explained, that it would be simple and brief [...] and easy to use. That it would be intuitive, very visual, that you could find everything you need or what the app offers, instantaneously.* [P02, aged 50 years, moderate PA level, neutral skill with technology]
>
> *If it [the app] wouldn't imply more work from me [...]. If you could understand well the different aspects it would help of course… that it would be friendly and useful.* [P09, aged 64 years, high PA level, neutral skill with technology]

Mixed Opinions on Playfulness

Interviewees had varied opinions on playful experiences as a motivational factor. A few recognized the potential positive effects of having game-like experiences, which were thought to bring fun into PA:

> *Wow, that [having a game-like experience] would be very good.* [P11, aged 69 years, moderate PA level, high skill with technology]
>
> *It should be friendly, useful and fun… otherwise, you get bored.* [P09, aged 64 years, high PA level, neutral skill with technology]

A variety of gamification elements, including points, rewards, levels, avatars, and competition between users, were discussed during the interviews. Participants suggested the option of

exchanging points for real rewards or prizes (eg, a session of yoga, swimming, or pilates) or that rewards could be set by the users themselves, and a few perceived receiving badges (eg, virtual medals) as fun. The implementation of different levels of *experience* seemed to be positively received by the interviewees. Some participants liked the idea of a virtual avatar that they could customize to look like them and to be able to see their activity progress through it. Finally, a few of them also believed that competitions with others could be interesting:

> *Well, if what you receive, the rewards, can be exchanged for something [...], the pilates session or anything like that...* [P14, aged 55 years, low PA level, high skill with technology]
>
> *And you've never thought about rewards being defined by each one? [...] I'd leave it a bit open, you know? [...] [For example,] I will reward myself going to a spa or a weekend getaway [...]* [P07, aged 56 years, moderate PA level, high skill with technology]
>
> *It is kind of a fun game, where I also imagine that you get challenged, right? You will want to level up... it seems good to me.* [P03, aged 48 years, moderate PA level, high skill with technology]
>
> *I like the avatar a lot and also with the dimensions, not a thin avatar. I am now like this, you put me like this, and as we evolve, my little avatar will be losing weight together with me.* [P08, aged 47 years, moderate PA level, high skill with technology]
>
> *But the aspect of competitions with others or doing it with a social network, I think that is quite interesting [...] it always helps you a lot in a challenge or simply to encourage you [...].* [P05, aged 43 years, moderate PA level, high skill with technology]

However, mixed opinions prevailed with some participants, suggesting that these game-like elements were not motivating for them, as they felt more appropriate to younger people or for those who are into games or are more competitive:

> *But we are not children. Points do not motivate us.* [P04, aged 48 years, moderate PA level, high skill with technology]
>
> *I think it does have something fun. What I don't know if all people, depending on age, may like it more or less.* [P06, aged 50 years, moderate PA level, neutral skill with technology]
>
> *There are those who like the theme of games so much, but it is not my case, so I tell you that if they give you four suns or three moons for having achieved it... come on, that does not call my attention at all, I do not find it attractive.* [P13, aged 47 years, moderate PA level, high skill with technology]
>
> *To me, the competitions don't… they put me nervous, I don't like them.* [P12, aged 66 years, moderate PA level, high skill with technology]

It was also suggested that the game experience should be optional to suit every user:

> *I believe that this part should be optional you know? You can start the game or not start the game … in that way, what happens is that the competitive person is stimulated, but the person that is not competitive and that can have certain stress from this [game], does not have to enter in this with herself.* [P07, aged 56 years, moderate PA level, high skill with technology]

Mixed Opinions on Interacting With Other Users

Some participants were interested in the idea of having a kind of social network in the app that would allow them to have some type of contact with other users. They liked the possibility of sharing, comparing with others, and getting recognition by others. It was also suggested that there could be a common goal for groups of users:

> *But the aspect of competitions with others or doing it with a social network, I think that is quite interesting […] because you go with other people to do something and the company and sharing it with other people is essential, at least in my case. [...]* [P05, aged 43 years, moderate PA level, high skill with technology]
>
> *I would not mind that it would be connected to Facebook, well it would be a way to upload it to Facebook and say I have improved this much. Some little message that is made public. But yes, yes, the recognition of others is important.* [P03, aged 48 years, moderate PA level, high skill with technology]
>
> *[...] that there would be several people and they could connect [...]. That they would set the goal of this week to walk this much, then to see who achieves it or something like that. That several people are in that group and everyone goes for the common goal.* [P14, aged 55 years, low PA level, high skill with technology]

Some participants did not feel that such features would suit them and thought it should be optional:

> *That it exists, that's fine, there's people who this would suit, specifically, to me, it would not suit me… I don't know. [...] we are complicated people, it's not easy, so… [...] I believe it should give the option in case there is people who benefit from it, but to me personally it would be complicated [...] it repels me.* [P09, aged 64 years, high PA level, neutral skill with technology]

It was also suggested that it could depend on the current treatment stage:

> *Well, I think it depends on the moment, right? Because when you are alone when you are in the first period [after treatment] maybe the digital coach will suit you. When you're a little better, maybe the social network and when you're pretty good, competition with others.* [P06, aged 50 years, moderate PA level, neutral skill with technology]

Personalization Aspects of the PA Coaching Experience

This theme highlights the participants' attitudes toward personalization and their perspectives on aspects that should be considered for creating an individualized PA coaching experience: targeting user characteristics, individualized progress information, dynamic adjustment of training, considering the user's situational context, and interface simulating a virtual coach.

Attitudes Toward Personalization

The participants had very positive feelings toward the idea of having an app experience individualized to each user. They often associated tailoring, or personalization, to an increase in user satisfaction or an increase in user engagement. Participants highlighted its particular relevance to address the individual needs of each breast cancer survivor:

> *Yes, totally, because the more it suits the interests of someone the more satisfied the person will be.* [P07, aged 56 years, moderate PA level, high skill with technology]

> *I think the idea that it is something personalized is very important. [...], the fact that it collects a lot of data about you and the circumstances of each one and some symptoms and other stuff like that… it seems to me to be the most important in order to create an app if it is different from what there is now that is not that personalized. [...]* [P05, aged 43 years, moderate PA level, high skill with technology]

> *Personalized, because they would be treating me specifically. Every person is unique and also every disease.* [P13, aged 47 years, moderate PA level, high skill with technology]

Some highlighted its importance, particularly in the early stages of their cancer journey:

> *During the whole process [of the disease and treatment] it [personalization] seems very important to me. Maybe later, not so much, but in those moments, very much. [...]. [...] during the first moments I think it is fundamental [...].* [P06, aged 50 years, moderate PA level, neutral skill with technology]

Targeting User Characteristics

Participants often expressed the need to have an app tailored to their condition as breast cancer survivors but also on an individual level. Several factors were identified to be considered for personalization, including general characteristics (age, gender, weight), physical and emotional status (eg, limitations, pain, and stress), treatment stage and side effects from treatment and medication, and personal goals:

> *It would be ideal, because then if it is personalized... if that application knows my limitations, or whatever, or if I can add my limitations... the exercises would aim to meet my needs.* [P01, aged 57 years, low PA level, neutral skill with technology]

> *Personalized in terms of the side effects produced by a type of illness and medication. […] Maybe, a degree of pain or a degree of ... or a concept of nutrition or to free you of stress. [...] if there is an adjustment for stress or for pain or for whatever... a little more personalized, that's fabulous, of course.* [P02, aged 50 years, moderate PA level, neutral skill with technology]

> *I imagine it adjusted to each person. I cannot have the same guidelines as the person who has participated previously... there has to be an assessment of everything, the age, what type of specific illness you have had, what you want to achieve with what you are doing. I think that all the factors that should be considered, in order to personalize it a little more.* [P08, aged 47 years, moderate PA level, high skill with technology]

In addition, the user's preferences were considered as a factor that could be taken into account by the app. For example, it was suggested that the app could consider the user's activity preferences in the communication provided or in the activity program:

> *It's good because then it reminds you of what [activities] you like. Maybe I'm having a day that I'm feeling down, and it reminds me of what [activity] I really like. I like to be in my garden [...]. [So,] I go to my garden and I start doing this [garden activity] that is good for me.* [P07, aged 56 years, moderate PA level, high skill with technology]

> *[For example,] "So, look… running doesn't suit you. You could walk for one hour and a half […]".* [P08, aged 47 years, moderate PA level, high skill with technology]

It was also suggested that some of the app functionalities could be optional to suit one's preferences:

> *Always that things [functionalities] are optional, it seems ok to me… if something doesn't suit me, it doesn't mean it will not suit another person.* [P07, aged 56 years, moderate PA level, high skill with technology]

Individualized Progress Information

Participants suggested that it would be good to get feedback on their current progress toward goal, particularly when they were close to achieving it, as it would work as an incentive to complete it:

> *[Talking about a commercial app] It tells you "you are very close to your objective today"... that's very good because it helps you say "ok, I'm going to do it [the activity] for another little while".* [P02, aged 50 years, moderate PA level, neutral skill with technology]

In addition, they wanted to access data on their past activity achievements to understand how they had progressed over time:

> *Something like that [talking about a game that showed progress in graphs], weekly or every fifteen days, I don't know, that it gives you progress in various formats. Weekly, monthly, at the end of the year and*

> *you see how you overcome obstacles.* [P06, aged 50 years, moderate PA level, neutral skill with technology]
>
> *It's like now if you don't write down how much you walk every day, in the end, they ask you within a month and you don't know if in the second week I went to walk every day or one day... so I see it very well, for the aspect of keeping track of your physical activity.* [P13, aged 47 years, moderate PA level, high skill with technology]

Dynamic Adjustment of Training

Participants wanted the app to have activity goals adjusted in difficulty to their PA level and progress. This need for goal adaptation is associated with the competence support characteristics of dynamic difficulty and growth adaptation from SDT. It was also suggested that the app could challenge the user by, for example, increasing the proposed distance or speed when walking. This is associated with the concept of appropriate challenge and characteristics of competence support:

> *That the person who undertakes them [objectives] finds them easy to achieve and that, as we spoke before, these objectives would vary if you have been accomplishing part of them… that they would be each time more complex and allowing you to overcome yourself.* [P02, aged 50 years, moderate PA level, neutral skill with technology]
>
> *That the application would recommend you or dare you to achieve, I don't know, for example when walking, [to do] more kilometers or maybe in less time, so that you increase your speed… or something like that. It would be interesting. Especially, because it is true that one gets bored when doing certain things... and you stay a little stuck.* [P05, aged 43 years, moderate PA level, high skill with technology]

Participants expressed a desire for the training to be adjusted to how the user was feeling and to the treatment stage:

> *That it can register everything that you are going through in that moment to adjust as much as possible the subject of training, you know? To me, it seems super important. [...] For example, the message of "if you are tired walk every five minutes".* [P05, aged 43 years, moderate PA level, high skill with technology]
>
> *If you are at the beginning of treatment […] if it warns you with a reminder to tell you "today you have already done enough". That when you have already recovered it is a reason to encourage you… but that it would be also a reason to stop you or to force you to pause [...].* [P06, aged 50 years, moderate PA level, neutral skill with technology]

Considering the User's Situational Context

Situational context is information that can be used to characterize the situation of a user (not including the user's characteristics). Most participants preferred outdoor activities and highlighted the importance of weather conditions as an influencing factor for performing PA:

> *The weather is important because what I do is walking. If it is very hot or very cold I do not go out [...]. This month has been very cold and I have not gone out.* [P01, aged 57 years, low PA level, neutral skill with technology]

There was an interest in having an app that would be aware of their location to provide relevant recommendations. It was suggested that it could suggest nearby places to perform an activity, provide alternative routes, and provide alternative indoor exercises for rainy days:

> *[...] everything about the weather and the location [...] it seems good. That it also informs you about what there is around you, what may interest you… it seems interesting to me.* [P03, aged 48 years, moderate PA level, high skill with technology]
>
> *About routes and getting alternative when maybe the weather is bad or whatever, I find it very interesting. […] I think it is interesting if you can get a training program, maybe for doing at home or something like that.* [P05, aged 43 years, moderate PA level, high skill with technology]
>
> *[...] I always go through the same place, then it can suggest other routes.* [P14, aged 55 years, low PA level, high skill with technology]

In a different aspect of situational context, some participants were also interested in having their sedentary behavior tracked, to have reminders or warnings to avoid such behavior, and recommendations to move more:

> *If you get an alarm, like "you've been sitting for two hours I recommend you to move" or something else, yes, it would be good. [...].* [P03, aged 48 years, moderate PA level, high skill with technology]

Interface Simulating a Virtual Coach

Participants seemed to like the idea of having a virtual coach, simulating the interactions with a real trainer. They suggested that it was an interesting and entertaining way of providing more personal feedback and guidance on PA:

> *I imagine it as a coach in a minicomputer, that would be ideal. […] Yes it would be nice, it would be curious to be able to interact...* [P08, aged 47 years, moderate PA level, high skill with technology]
>
> *[...] also, the personal assistant is entertaining and friendly and it gives you the sensation of personalization.* [P09, aged 64 years, high PA level, neutral skill with technology]
>
> *That seems very interesting to me, I think it's great because it also makes the application more entertaining, brings it closer to the person, at least to me. [...] Yes... maybe the theme of a virtual coach makes it more enjoyable, no? [...] And obviously it is very good because it's that, it's a control and it's something informative.* [P13, aged 47 years, moderate PA level, high skill with technology]

It was also suggested that the virtual coach could bring a feeling of trust:

> *[...] maybe the virtual assistant, or creating a chat, make it like, let's say, like more human, right? And it would give you, perhaps, more confidence.* [P13, aged 47 years, moderate PA level, high skill with technology]

Technology Trustworthiness

This theme gathers insights from participants on system requirements related to trusting the content and with the use of data collected from the app. The subthemes addressed are system and content credibility and data sharing and privacy.

System and Content Credibility

Participants highlighted the importance of trusting the information they are provided with. Some expressed their concern about assessing which information found on the internet was appropriate for them:

> *There are things that I read that are then generic and are contraindicated for breast cancer issues, so that causes me a lot of insecurity in webs… I would like a serious app where I do not put at risk what is really happening to me. Something that does not put my health at risk.* [P03, aged 48 years, moderate PA level, high skill with technology]

In addition, some suggested having more trust in information and tools if they were recommended or validated by their health care professionals or backed by exercise specialists. This is associated with the concept of verbal persuasion from SCT:

> *[...] it would be very interesting to see that there are specialists behind, that is guided by people who know what they propose and who know what they have in hand, like: "we are coaches", "we are athletes", "we are sportspeople". This is very important of course. It's what gives you security in what you are doing...* [P02, aged 50 years, moderate PA level, neutral skill with technology]

It was suggested that finding a contact from someone involved in the development of the app or finding that there were other users of the app would make it feel more real:

> *Well, that's true, that whenever I enter a page there is always some contact, that there is a telephone, an address because it makes it more real, that, although you are on the web, at any given time you can get in touch with a person, like in a more real way.* [P13, aged 47 years, moderate PA level, high skill with technology]

> *[If] you can connect through the page with people that are in the same circumstances [...], it makes it more real.* [P13, aged 47 years, moderate PA level, high skill with technology]

Data Sharing and Privacy

Participants showed a positive attitude toward extensive data collection for the purpose of personalizing the experience as long as their privacy was secured. Some highlighted that they would only do it if the tradeoff between personal data sharing and experience benefits was positive for them:

> *If the benefit for me is greater than the loss of privacy with data management, well I would do it [share their data].* [P06, aged 50 years, moderate PA level, neutral skill with technology]

> *The more data, the more personalized and that seems good to me. What I can get insecure about is giving away that data so [...] the application should be safe, through codes, passwords. Yes, because I think that if it's personalized it's much better.* [P13, aged 47 years, moderate PA level, high skill with technology]

However, they expressed their concerns toward data sharing and access by others. Participants agreed to share data with professionals, whereas preferred it to be optional with other users. In general, they expressed their wish to be able to decide who they share their data with:

> *With medical personnel, with personnel who I know will handle that data, not with anyone. And besides, if you share these data with a medical purpose of improving the application or to help other people… I don't know, it seems also important to me when it comes to assessing the loss of intimacy.* [P06, aged 50 years, moderate PA level, neutral skill with technology]

> *[...] within the community I would not want to share my data with anyone unless I expressly say that I want to share […]. Within the internet community there are many bad people hidden in a profile like this...* [P07, aged 56 years, moderate PA level, high skill with technology]

> *Well, I'd like it to have the option of who can see and who can't.* [P06, aged 50 years, moderate PA level, neutral skill with technology]

Discussion

Principal Findings

In this study, we conducted semistructured interviews with breast cancer survivors to identify specific needs and considerations regarding various aspects associated with motivation and personalization in PA coaching apps. We identified 5 overarching themes that can guide the design of a future solution: (1) barriers to PA, (2) psychological mediators of PA motivation, (3) needs and suggestions for reinforcing motivation support, (4) personalization aspects of the PA coaching experience, and (5) technology trustworthiness. Important findings from this study include the identification of various determinants of motivation associated with PA adherence: perceived control of behavior and goals, confidence and perceived growth, belief in PA outcomes, and social connectedness. A variety of intervention needs and suggestions for reinforcing motivation were also identified: activity monitoring and goal setting, PA prescription, positive reinforcement, absence of pressure, simplicity and ease of use, and mixed opinions on playfulness and on interacting with other users. Importantly, considerations for personalization of the coaching experience were reported: attitudes toward personalization, targeting user characteristics, individualized progress information, dynamic adjustment of training,

considering user's situational context, and interface simulating a virtual coach. In addition, potential barriers for PA adherence were identified, which helped contextualize some of the motivational and personalization aspects, including physical and emotional challenges, lack of time, and lack of information. Finally, aspects of technology trustworthiness were also highlighted regarding system and content credibility and data sharing and privacy. Overall, survivors believed that a simple and personalized app, which addresses their individual needs and preferences and that provides feedback, guidance, and encouragement for the achievement of progressive PA goals, could be helpful in engaging in a more active lifestyle.

Altogether, these findings can inform the design and may help increase the acceptability and sustained interest of theory-based mHealth PA interventions for breast cancer survivors.

Comparison With Previous Work

Barriers for PA

The findings reported in this study indicate that breast cancer survivors often face barriers to PA participation, which may affect their well-being [16,45]. The 4 barriers described by the participants in this study are somewhat connected and can have a ripple effect, as suggested in the literature [46]. Lack of time and inconvenient exercise schedules are usually reported as one of breast cancer survivors' bigger barriers to PA [47] that can be a consequence of professional work hours and family caregiving roles [16,45]. Cancer-related fatigue, pain, and weight gain were at the center of physical barriers experienced by the participants and could interfere with the daily functioning of the patients along with their quality of life [16,45,46,48]. Emotional problems (eg, not feeling good or stress) were also reported as a PA barrier, and seem to be a consequence of the psychological burden associated with sticking with the long treatment regimens, which may last for the rest of their lives [16,45,48]. Finally, lack of information was also suggested as a key barrier. When patients with breast cancer cannot access proper information, they can end up following wrong and unguided routines, which may cause frustration or even physical strain and, ultimately, make people give up from continuing with PA [16].

Other barriers reported in the literature that were not identified in our results include access to facilities, aging process, seasonal weather, and disliking exercise [16,45,46,48]. However, some participants in our study did suggest that age and weather aspects should be considered as factors for personalizing the coaching experience.

Opinions and Preferences for mHealth PA Interventions

In their study, Phillips et al [14] conducted a mixed-methods study to identify the preferences of breast cancer survivors for mHealth PA intervention features. The themes identified from their interviews were importance of relevance to breast cancer survivors, ease of use, integration with wearable activity trackers, providing a sense of accomplishment, and variability in the desired level of structure and personalization. In their quantitative study, participants revealed a preference for daily and weekly progress feedback, newsfeed, activity challenges, scheduling tools, and motivational and reminder messages. The studies are in line with our study, which also demonstrated an interest of the participants for such intervention components. Integration with wearables was not highlighted in our data, but could be used in combination with a mobile PA coaching app to facilitate activity monitoring and avoid the possible burden of carrying a smartphone during PA. In this sense, the system could leverage the unobtrusive nature of wearables during activity when keeping the richer coaching capabilities from smartphone apps [12].

Before that study, Phillips et al [18] conducted quantitative research through web-based questionnaires to explore breast cancer survivors' interests and preferences for technology-supported exercise interventions. Their findings are consistent with ours in that the majority of their participants reported that personalized feedback was seen as one of the most helpful technology intervention components. In addition, the least rated components by participants were social networking, group competitions, and the ability to see others' progress, which relate to our findings about the mixed opinions on interacting with other users.

In the formative development of the Bounce app, a mobile app to increase PA in breast cancer survivors, the research team developed a framework and presented a set of guidelines for the design of behavioral intervention technologies for breast cancer survivors [19]. However, given the focus on reporting intervention design decisions based on the proposed framework, their findings from patient interviews were not detailed in the paper. In addition, they considered a combination of SCT and TTM, whereas our study used the SCT and SDT perspectives. Despite this, their empirical findings seemed to be mostly in line with our results, indicating similar constructs, including reinforcements, helping relationships, self-monitoring, goal setting, verbal persuasion, and mastery experience. In terms of the intervention components, the studies are consistent on the possible inclusion of a scheduling tool; progress monitoring tools, with visuals on current activity (distance and time) and weekly summaries; and a progressive activity program, with incremental levels adjusted to the user's progress. However, their study simply identified the need for opportunities to connect with other breast cancer survivors in a similar context, whereas in our findings, the social networking aspects did not seem to suit everyone. In addition, owing to the difference in nature of the types of activity considered in the studies, some components were different. For example, the Bounce app was designed to include flexibility and strength exercises, which led to requirements associated with reassuring safety (eg, video demonstrations).

In the most recent paper about the Bounce app [24], gamification techniques such as badges and trophies to reward the users and motivational themes for data visualization were explored. However, with regard to the gamified themes, the authors suggest that they have identified the overall preference of users toward more straightforward representations of their numerical data. This, in some way, is in line with our findings on playfulness, as the variability in opinions about gamification prevailed, with some participants finding its potential to create fun and friendly experiences, although others saw it more appropriate for younger or more competitive people.

Our study complements the findings from previous work with qualitative research on breast cancer survivors' perspectives on motivational aspects and personalization strategies associated with PA intervention adherence. The findings provide a level of detail across psychological, intervention, and individualization needs for mobile PA coaching that, to the best of our knowledge, was not performed in previous studies on this topic.

Overall, studies including ours demonstrate that breast cancer survivors believe that a mobile tool for PA guidance and coaching can be helpful to adhere to the PA recommendations and are interested in a variety of potential features.

Compared with other intervention modalities, smartphones have a portable size and can use built-in sensors or be paired with external sensors for continuous activity monitoring and feedback. In addition, smartphones' screen sizes, despite being smaller than tablets or portable computers, also allow for creating rich app experiences to the users [12]. Hence, these devices are believed to be the most suitable for real time PA coaching systems. Despite this, it is cautioned that low technology literacy may be a barrier to adherence in some cancer survivors [13,14] and that apps should be designed to be user friendly and not overly complex. However, survivors are typically older adults, and technology use in this segment of the population is increasing rapidly [49]. Future quantitative research should address the feasibility and efficacy of these intervention modalities and specific features in PA coaching systems for this population and consider a high range of digital literacy levels.

Recommendations for Research and Practice

The findings presented in this study can be explored to design more motivating and personalized mHealth PA coaching interventions for breast cancer survivors. The insights provided help to understand how to satisfy the psychological needs associated with PA adherence and to increase perceived personal relevance and engagement, which in turn are thought to create value, develop intrinsic motivations, and consequently, lead to sustained adherence. A variety of key recommendations and considerations in this direction are provided in Textboxes 1 and 2.

Our findings suggest that key behavioral determinants of PA adherence in breast cancer survivors can be seen through the lens of SCT and SDT and be used to reason and inform the construction of future interventions. For example, the theme of perceived control of behavior and goals was identified as a motivational determinant associated with PA adherence and is related to the component of autonomy from SDT. Given that SDT posits that intervention characteristics such as increase in sense of ownership, absence of pressure, and self-regulation can support perceived autonomy [31,44], an app could be designed to accomplish these by, for example, personalizing and customizing content to the user, not sending negative feedback to users, and providing opportunities for tracking daily activity. Similarly, to help inform such design decisions, the results section analyses the relationship between the suggested intervention needs (eg, progress feedback, PA prescription, positive reinforcement) and the identified psychological determinants of PA motivation (eg, autonomy, self-efficacy). Finally, matching these insights from the users and theory with, for example, the Coventry, Aberdeen, and London-Refined taxonomy of behavior change techniques [50] can help researchers refine or add new components for the design of an app aimed at increasing PA. Textbox 1 presents a number of considerations regarding the integration of behavior change methods based on such taxonomy and on the positive computing book [44].

Despite the mixed reactions toward more social and gamified functionalities, these may have a critical motivational role for some users and should, therefore, be carefully considered in the design, made optional to the user, or have a secondary role in the app. In particular, based on the participants' perspectives, it seems worth exploring strategies for connecting with similar users, sharing the experience with loved ones, and getting support from a psycho-oncologist. In addition, the use of game-like mechanisms driven toward an increase in intrinsic motivation (eg, including levels and setting rewards that are valued by users and provide a sense of progress) and that do not make the experience childish or very competitive may contribute to engaging the users.

The variability in opinions regarding certain app characteristics also highlights the need for an individual experience and the importance of personalizing content to the users' preferences. As suggested by Phillips et al [14], mHealth solutions provide a unique opportunity to create highly personalized interventions for breast cancer survivors in real time. The participants in our study considered the following factors essential for individualization: general characteristics (eg, age, gender, weight), physical status (eg, limitations, fatigue, pain), emotional status (eg, general feelings, stress, and anxiety), treatment stage and side effects, personal goals, preferences, PA level and progress, and weather and location. In addition, participants provided a variety of suggestions for a more individualized experience, which align with the model and strategies for real time personalization in PA coaching systems [12,20]. Associating our findings with these strategies can help design a personalized coaching experience for these individuals. For example, the content of feedback can be explored to achieve user-targeted communication by simply including the user's name in the messages or by considering the user's preferences and personal goals to present information specific to the user's interest (eg, if a user's main goal is weight management, the app could display more specific feedback information about caloric data). In addition, an app may set a training plan and goals based on the user characteristics (eg, age, baseline PA level, and symptoms) and dynamically adjust these through time according to their progress and how they are feeling (eg, mood, stress, pain). More complex forms of personalization, such as self-learning, may also be explored to select opportune moments for the delivery of activity cues or to adjust training difficulty based on system usage. Suggestions for how an app might incorporate these and other personalization strategies are provided in Textbox 2 on the basis of the model for real time tailoring in PA coaching systems [12,20].

Textbox 1. Key behavior change recommendations for the design and development of mobile health physical activity coaching apps and intervention methods (on the basis of the positive computing book and the Coventry, Aberdeen, and London-Refined taxonomy) for breast cancer survivors.

Sense of ownership; ability to customize

- App should align with the user's own value-based goals (eg, a user may want to have a more active lifestyle to lose weight, prevent recurrence of disease, or be more fit to be able to carry on with their roles in work and family) in the communication and training provided
- Allow for a certain degree of customization by the user (eg, goals, frequency, timing and type of activity cues [reminders or notifications], and types of activity)

Clear rationale

- Provide a clear explanation of the reasoning behind the app, coaching, training program, and goals; this may include a summary of the evidence in lay terms and may target the user's own values and goals

Self-regulation, self-monitoring, goal setting

- Set physical activity (PA)–related short-term goals (eg, on a daily and weekly basis) and encourage users to set long-term goals and expected outcomes that are aligned with user's own values
- Monitor and present visual information on current activity (eg, steps, calories, pace, distance, and duration), progress toward goals, and progress through time (eg, weekly, monthly, and yearly historic data)
- Consider monitoring sedentary behavior (eg, time spent sitting)
- Provide interpretation of overall progress (eg, weekly and monthly) relevant to user's values and concerns
- Consider integrating wearables to facilitate activity tracking and real time feedback

Mastering new skills, dynamic difficulty, appropriate challenges

- Have users starting with goals that are fairly comfortable to accomplish and increase goals and training difficulty over time
- Set users with activity challenges to put their fitness level and skill to test
- Adjust training difficulty and challenges to users performance level

Absence of pressure

- Just enough reminders and notifications
- Avoid negative feedback when goals are not reached; instead, cheer up users and encourage them to come back to it another day
- Allow for periods of rest or periods with light training, when needed

Positive reinforcement, verbal persuasion, constructive feedback

- Positive and casual tone of communication from a trusted source (eg, create feedback messages together with professionals)
- Use clear and easy to understand messages
- Acknowledge progress and achievements; possibility to explore rewards based on effort and that are relevant to the user
- Highlight that the sustained PA achievements may bring relevant health benefits

Assist in time management, use follow-up prompts

- Include a tool for planning and scheduling activity (eg, on a weekly basis)
- Include periodic reminders for activity
- Scheduling and reminders should be customizable by the user (eg, allow users to reschedule activities)

Prompt self-talk

- Encourage users to use positive self-talk for motivation to comply with the plan and during activity sessions

Stimulate anticipation of future rewards

- Encourage users to think about the positive outcomes of achieving their PA objectives on their condition as breast cancer survivors (eg, prevention of cancer recurrence and better quality of life) and on their value-based goals (eg, being more fit to play with their children or grandchildren)

Instructed practice

- Provide a prescription of PA with detailed activities, goals, and explaining or demonstrating proper technique
- Provide feedback on user's performance of the activities

- Address user's physical limitations

Stress management

- Consider including guidance on how to cope with emotional challenges
- Suggestions for relaxation and meditation exercises could be included
- Encourage connection with a psycho-oncologist

Playfulness

- Gamification elements could be explored for providing fun and engaging experiences, but should be carefully designed, made optional to the user, or have a secondary role in the app; avoid creating a gamified design (eg, systems for points or rewards) that may be perceived as childish or very competitive
- Points, progress bars, levels, and challenges may be considered to provide a sense of progress to the user; customizable avatars may also be explored in this way by, for example, making them evolve physically along with user's progress
- Explore rewards that may be exchanged by real things and experiences (eg, sessions of yoga or Pilates); consider using rewards that are valued by the user or that can be in some way customized by them
- Optional *healthy* competitions, activity comparisons, or goal sharing within a user community may also be considered in the design; consider user's personality in terms of competitiveness and openness to social aspects

Provide opportunities for social support

- Optional feature for sharing experience (eg, achievements, progress) with close ones
- Optional networking feature to connect or to share goals with other similar users (eg, buddy system or feature to team up with others to achieve community goals)
- Consider including feature to allow for connection with a counselor (eg, psycho-oncologist) or an exercise trainer for support

High usability

- Make app easy to operate
- Include straightforward and simple content and interface
- Include instructions on how to use the app
- Include explanations on interface specifics (eg, explanation on what a graph represents) and specific action instructions (eg, "Type in your name in the box below")

Tailoring (or personalization)

- Provide tailored coaching experiences relevant to the user; tailor to the users on a group level, as breast cancer survivors, and on an individual level
- Individualization factors to consider include general characteristics (eg, age, gender, weight), physical status (eg, limitations, fatigue, pain), emotional status (eg, general feelings, stress, and anxiety), treatment stage and side effects, personal goals, preferences, PA level and progress, and weather and location
- Maximize automatic forms of tailoring to reduce need for data input by the user
- Consider the model of real time tailoring in PA coaching apps and the recommendations for personalization provided in Textbox 2

Provide a sense of trust and privacy; verbal persuasion

- Consider including contact details of the people involved in the app development
- Develop based on insights from breast cancer survivors and, together with experts in the field, take an evidence-based approach and validate with health care professionals (eg, oncologists and experts in PA)
- Make these steps explicit to the user
- Inform the user of the total number of breast cancer survivors using the app
- Provide clear and easy to understand information on data security and privacy methods used
- Include transparent and customizable data sharing (eg, with professionals, researchers, other users)

Textbox 2. Key personalization recommendations for the design and development of mobile health physical activity coaching apps and interventions for breast cancer survivors.

Feedback

- Present the user with information about their own activity data and progress; use straightforward and visual forms for representing data (eg, clear and succinct text, graphs, and progress bars)
- Consider having an optional feature to present feedback in comparison with other similar users (feedback associated with user targeting and interhuman interaction)
- Communication could be represented as a simulation of a virtual coach

User targeting

- Convey that the communication is designed specifically for them (eg, a message in the app could start with "Based on your activity level and preferences…")
- Include the user's name in the communication
- Present activity and progress information based on the user's likes (eg, user may like to see steps and distance but not calories)
- Present activity suggestions, goals, and information based on the user's characteristics (eg, age, weight, physical activity [PA] level, physical limitations, and likes)
- Present information and tips considering user's physical limitations

Goal setting

- Dynamically adjust the difficulty of the training plan or goals based on user's initial PA level, progress, how they are feeling (eg, mood, stress, and pain), and the perceived difficulty
- Suggest resting periods based on user's progress and how they are feeling

Interhuman interaction

- Explore optional networking or buddy system features joining only similar users based on user characteristics, PA level, and cancer experience (eg, cancer type, time since treatment, and side effects)
- App could include interactions with a real human coach

Adaptation

- Motivate users to achieve their goals and target progress information on the user's outcome expectations (eg, prevention of cancer recurrence, increased physical capability, weight loss, and increased satisfaction)
- Depending on the stage they are at in their cancer experience, suggest or highlight different features (eg, in the first stages posttreatment it might be important to have support from a counselor, whereas connecting with other users may only be accepted in later stages)
- Adjust type of communication based on user's confidence levels and skills. For example, if user has low confidence and skills, the app could provide more verbal persuasion and acknowledgement and then adjust it through time based on the user's growth
- Assess user's personality profile to define set of features that are relevant to the user; consider this particularly for more social and gamified features (eg, if the personality traits of a user suggest that the user is socially open and competitive then the app would adapt to include components for networking and competing with other users)

Context awareness

- Consider the user's location to provide suggestions for outdoor activities and alternative routes
- Encourage users to keep doing PA when traveling with suggestions of fitness facilities or outdoor places to do PA
- Encourage users to do PA on bad weather days and provide suggestions for indoor activities
- Adjust goal setting based on user's working schedules (eg, if a user works full time during the week then set goals for outside working hours or allow to make up for it on the weekends)

Self-learning

- An app may select opportune moments for delivery of activity cues (eg, reminders, notifications) or suggest new training schedules based on system usage (previous interactions of the user with the system)
- Self-learning can be used in combination with goal setting to infer the user's progress through time and adjust the training difficulty appropriately
- Self-learning strategies may also be used in combination with adaptation to automatically adjust the user's model based on their stage in the cancer experience or competence skills to provide adapted content or features to the user
-

An app may track user's response to the feedback and recommendations provided by the system, learn their preferences, and adjust the recommendations appropriately

More personalization also means more data being collected from the user. Hence, data sharing and privacy requirements need to be understood and considered in the design of these solutions. All participants indicated a willingness to share their personal data with the app in exchange for a more personalized experience as long as they could choose which data to share and who accesses the data and whether security would be guaranteed. In line with this, it seems necessary that personal data sharing is optional and transparent, and that users trust the app. Technology trustworthiness may be increased by informing users that the system was created together with health care and exercise professionals, and that there are other individuals like them using it. In addition, it is believed that better system credibility may increase the persuasive capability of these PA-promoting apps [51].

Future work exploring the user-centered design of a mobile PA intervention for breast cancer survivors should take into consideration the insights provided in this and related studies, explore how to best integrate the motivational and personalization strategies suggested here, involve the end users in the conceptualization and evaluation processes, and be informed and validated by professionals. In addition, future research should be conducted to assess the impact of these different features and combinations of features on breast cancer survivors' engagement with these systems.

Limitations

The results should be interpreted in the context of its limitations. All participants were recruited from the same oncologic clinic, meaning they had received similar care for their cancer. Our sample may be more engaged with the topic owing to their degree of awareness and moderate participation in PA according to the IPAQ-SF. In addition, all the participants were White, Spanish, and posttreatment breast cancer survivors, and most were educated and had access to technology. Despite their awareness of the importance of PA and their moderate IPAQ-SF levels, almost 60% of the participants did not adhere to the recommended levels in the guidelines, and most reported about the lack of detailed information on PA. In addition, although most participants were highly educated and reported high access and usage of technology, some had a neutral self-reported skill with technology or suggested having low digital literacy during the interviews. Future work should generalize to a more diverse sample of breast cancer survivors, considering age, employment status, received cancer care, educational level, digital literacy, country, and race or ethnicity. In addition, it could be argued that less-active participants may need a substantively different intervention approach. Therefore, future studies should analyze such differences and explore the stages of change of the TTM and its constructs, which may provide useful insights in such a direction. Finally, using qualitative research software such as NVivo could have assisted in the thematic analysis and provide more comprehensive insights into the data.

Conclusions

This work identifies a number of motivational and personalization factors and strategies to be explored in the design of PA coaching systems for breast cancer survivors from the end users' perspectives. It was grounded in relevant behavior change theories and techniques and the model of real time personalization, which are believed to help create successful interventions. Overall, the findings suggest the need to develop simple, guiding, encouraging, and trustworthy PA apps personalized to breast cancer survivors at both the group and individual levels. This paper opens up new possibilities for the design of PA coaching experiences for these individuals, which may ultimately help sustain technology adherence and increase PA participation.

Future studies should incorporate the perspectives of health care, sports science, and technical professionals as well as further investigate these findings with the involvement of breast cancer survivors in the design, testing, and implementation of PA app prototypes.

Acknowledgments

This study received funding from the European Union's Horizon 2020 research and innovation program: Marie Sklodowska-Curie Actions grant agreement no. 722012-CATCH-Cancer: Activating Technology for Connected Health. The publication was funded by Cátedra Telefónica "Inteligencia en la Red" of the University of Seville. EDZ is supported by the "V Plan Propio de Investigación" of the Universidad de Sevilla. This study was performed in collaboration with Salumedia Tecnologias SL, the Oncoavanze Clinic, the University of Seville, and the Insight Centre for Data Analytics from the University College Dublin. The authors would like to thank Dr Pedro Valero, Medical Director and Oncologist at Oncoavanze, all the Oncoavanze personnel, and all the patients who participated in the studies for their involvement and interest in this research. All authors acknowledge the valuable feedback provided by Darragh McCashin on earlier drafts.

Authors' Contributions

FG led this qualitative study, contributed to the analysis, and wrote the main body of the manuscript. GS, OR, and EZ contributed to the analysis and supported the manuscript. BC contributed to the study conception, design, organization, and revision of the manuscript.

Conflicts of Interest

FG and GS worked for at least some period of the study at Salumedia Labs, a digital health company that develops and commercializes mHealth solutions for supporting cancer patients. The remaining authors declare no conflict of interest.

Multimedia Appendix 1
Interview guide for the first part of the sessions.
[DOCX File , 15 KB - mhealth_v8i9e18867_app1.docx]

Multimedia Appendix 2
Slideshow presentation of features for the second part of the sessions.
[PPTX File , 33011 KB - mhealth_v8i9e18867_app2.pptx]

Multimedia Appendix 3
Participants' access to technology and technology usage.
[DOCX File , 945 KB - mhealth_v8i9e18867_app3.docx]

Multimedia Appendix 4
Supporting quotation for 5 analytical themes and their descriptive subthemes.
[DOCX File , 37 KB - mhealth_v8i9e18867_app4.docx]

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55(2):74-108 [FREE Full text] [doi: 10.3322/canjclin.55.2.74] [Medline: 15761078]
2. Sun Y, Zhao Z, Yang Z, Xu F, Lu H, Zhu Z, et al. Risk factors and preventions of breast cancer. Int J Biol Sci 2017;13(11):1387-1397 [FREE Full text] [doi: 10.7150/ijbs.21635] [Medline: 29209143]
3. Weingart SN, Brown E, Bach PB, Eng K, Johnson SA, Kuzel TM, et al. NCCN task force report: oral chemotherapy. J Natl Compr Canc Netw 2008 Mar;6(Suppl 3):S1-14. [Medline: 18377852]
4. Cho J, Jung S, Lee JE, Shim E, Kim NH, Kim Z, et al. A review of breast cancer survivorship issues from survivors' perspectives. J Breast Cancer 2014 Sep;17(3):189-199 [FREE Full text] [doi: 10.4048/jbc.2014.17.3.189] [Medline: 25320616]
5. Schmitz KH, Speck RM. Risks and benefits of physical activity among breast cancer survivors who have completed treatment. Womens Health (Lond) 2010 Mar;6(2):221-238. [doi: 10.2217/whe.10.1] [Medline: 20187728]
6. Duijts SF, Faber MM, Oldenburg HS, van Beurden M, Aaronson NK. Effectiveness of behavioral techniques and physical exercise on psychosocial functioning and health-related quality of life in breast cancer patients and survivors--a meta-analysis. Psychooncology 2011 Feb;20(2):115-126. [doi: 10.1002/pon.1728] [Medline: 20336645]
7. Smith SG, Chagpar AB. Adherence to physical activity guidelines in breast cancer survivors. Am Surg 2010 Sep;76(9):962-965. [Medline: 20836343]
8. Mason C, Alfano CM, Smith AW, Wang C, Neuhouser ML, Duggan C, et al. Long-term physical activity trends in breast cancer survivors. Cancer Epidemiol Biomarkers Prev 2013 Jun;22(6):1153-1161 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-0141] [Medline: 23576689]
9. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
10. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: a systematic review. Health Informatics J 2016 Sep;22(3):451-469. [doi: 10.1177/1460458214567004] [Medline: 25649783]
11. Middelweerd A, Mollee JS, van der Wal CN, Brug J, te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
12. op den Akker H, Jones VM, Hermens HJ. Tailoring real-time physical activity coaching systems: a literature survey and model. User Model User-Adap Inter 2014 Jun 25;24(5):351-392. [doi: 10.1007/s11257-014-9146-y]
13. Robertson MC, Tsai E, Lyons EJ, Srinivasan S, Swartz MC, Baum ML, et al. Mobile health physical activity intervention preferences in cancer survivors: a qualitative study. JMIR Mhealth Uhealth 2017 Jan 24;5(1):e3 [FREE Full text] [doi: 10.2196/mhealth.6970] [Medline: 28119278]
14. Phillips SM, Courneya KS, Welch WA, Gavin KL, Cottrell A, Nielsen A, et al. Breast cancer survivors' preferences for mHealth physical activity interventions: findings from a mixed methods study. J Cancer Surviv 2019 Apr;13(2):292-305 [FREE Full text] [doi: 10.1007/s11764-019-00751-3] [Medline: 30912011]

15. Cabrita M, Op den Akker H, Tabak M, Hermens HJ, Vollenbroek-Hutten MM. Persuasive technology to support active and healthy ageing: an exploration of past, present, and future. J Biomed Inform 2018 Aug;84:17-30 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.010] [Medline: 29935348]
16. Hefferon K, Murphy H, McLeod J, Mutrie N, Campbell A. Understanding barriers to exercise implementation 5-year post-breast cancer diagnosis: a large-scale qualitative study. Health Educ Res 2013 Oct;28(5):843-856. [doi: 10.1093/her/cyt083] [Medline: 23969632]
17. 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]
18. Phillips SM, Conroy DE, Keadle SK, Pellegrini CA, Lloyd GR, Penedo FJ, et al. Breast cancer survivors' preferences for technology-supported exercise interventions. Support Care Cancer 2017 Oct;25(10):3243-3252 [FREE Full text] [doi: 10.1007/s00520-017-3735-3] [Medline: 28470368]
19. Plank M, Dalal N. Bounce: A Mobile Behavioral Intervention Technology for Breast Cancer Survivors. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA'16; May 7-12, 2016; San Jose, USA. [doi: 10.1145/2851581.2890383]
20. Monteiro-Guerra FM, Rivera-Romero O, Fernandez-Luque L, Caulfield B. Personalization in real-time physical activity coaching using mobile applications: a scoping review. IEEE J Biomed Health Inform 2020 Jun;24(6):1738-1751. [doi: 10.1109/JBHI.2019.2947243] [Medline: 31751254]
21. Schueller SM, Muñoz RF, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec 3;22(6):478-483. [doi: 10.1177/0963721413495872]
22. Milne H, Wallman K, Guilfoyle A, Gordon S, Corneya K. Self-determination theory and physical activity among breast cancer survivors. J Sport Exerc Psychol 2008 Feb;30(1):23-38. [doi: 10.1123/jsep.30.1.23] [Medline: 18369241]
23. Bluethmann SM, Bartholomew LK, Murphy CC, Vernon SW. Use of theory in behavior change interventions. Health Educ Behav 2017 Apr;44(2):245-253 [FREE Full text] [doi: 10.1177/1090198116647712] [Medline: 27226430]
24. Marcu G, Misra A, Caro K, Plank M, Leader A, Barsevick A. Bounce: Designing a Physical Activity Intervention for Breast Cancer Survivors. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PervasiveHealth'18; May 21-24, 2018; New York, NY, USA. [doi: 10.1145/3240925.3240927]
25. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
26. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2016 Jan;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]
27. Epstein DA, Caraway M, Johnston C, Ping A, Fogarty J, Munson SA. Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:1109-1113 [FREE Full text] [doi: 10.1145/2858036.2858045] [Medline: 28503678]
28. Lazar A, Koheler C, Tanenbaum J, Nguyen DH. Why We Use and Abandon Smart Devices. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2804288]
29. Ledger D, McCaffrey D. Inside Wearables - How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement. Medium. 2017. URL: https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2020-08-19]
30. Lee H, Uhm KE, Cheong IY, Yoo JS, Chung SH, Park YH, et al. Patient satisfaction with mobile health (mhealth) application for exercise intervention in breast cancer survivors. J Med Syst 2018 Nov 6;42(12):254. [doi: 10.1007/s10916-018-1096-1] [Medline: 30402781]
31. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
32. Deci EL, Ryan RM. Conceptualizations of intrinsic motivation and self-determination. In: Intrinsic Motivation and Self-Determination in Human Behavior. Perspectives in Social Psychology. Boston, MA: Springer; 1985.
33. Bandura A. Self-Efficacy: The Exercise of Control. New York, USA: W H Freeman; Nov 1, 1997.
34. Bandura A. Social Foundations of Thought and Action: A Social – Cognitive Theory. Englewood Cliffs, New Jersey: Prentice-Hall; 1986.
35. McAuley E, Blissmer B. Self-efficacy determinants and consequences of physical activity. Exerc Sport Sci Rev 2000 Apr;28(2):85-88. [Medline: 10902091]
36. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining 'Gamification'. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek'11; October 6-8, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
37. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]

38. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 4;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
39. 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]
40. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
41. Rock CL, Doyle C, Demark-Wahnefried W, Meyerhardt J, Courneya KS, Schwartz AL, et al. Nutrition and physical activity guidelines for cancer survivors. CA Cancer J Clin 2012;62(4):243-274 [FREE Full text] [doi: 10.3322/caac.21142] [Medline: 22539238]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Rhodes RE, Nigg CR. Advancing physical activity theory: a review and future directions. Exerc Sport Sci Rev 2011 Jul;39(3):113-119. [doi: 10.1097/JES.0b013e31821b94c8] [Medline: 21705861]
44. Calvo R, Peters D. Positive Computing: Technology for Wellbeing and Human Potential. Massachusetts, USA: MIT Press; 2014.
45. Lavallée JF, Abdin S, Faulkner J, Husted M. Barriers and facilitators to participating in physical activity for adults with breast cancer receiving adjuvant treatment: a qualitative metasynthesis. Psychooncology 2019 Mar;28(3):468-476. [doi: 10.1002/pon.4980] [Medline: 30657225]
46. Blaney J, Lowe-Strong A, Rankin J, Campbell A, Allen J, Gracey J. The cancer rehabilitation journey: barriers to and facilitators of exercise among patients with cancer-related fatigue. Phys Ther 2010 Aug;90(8):1135-1147. [doi: 10.2522/ptj.20090278] [Medline: 20558566]
47. Ventura EE, Ganz PA, Bower JE, Abascal L, Petersen L, Stanton AL, et al. Barriers to physical activity and healthy eating in young breast cancer survivors: modifiable risk factors and associations with body mass index. Breast Cancer Res Treat 2013 Nov;142(2):423-433 [FREE Full text] [doi: 10.1007/s10549-013-2749-x] [Medline: 24177756]
48. Cho D, Park C. Barriers to physical activity and healthy diet among breast cancer survivors: a multilevel perspective. Eur J Cancer Care (Engl) 2018 Jan;27(1):e12772 [FREE Full text] [doi: 10.1111/ecc.12772] [Medline: 28940854]
49. Smith A. Older Adults and Technology Use. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/2014/04/03/older-adults-and-technology-use/ [accessed 2020-08-19]
50. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
51. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]

Abbreviations

IPAQ-SF: International Physical Activity Questionnaire-Short Form
mHealth: mobile health
PA: physical activity
SCT: social cognitive theory
SDT: self-determination theory
SWLS: satisfaction with life scale
TTM: transtheoretical model

Edited by G Eysenbach; submitted 25.03.20; peer-reviewed by R Vromans, KT Win, M Robertson; comments to author 16.04.20; revised version received 22.05.20; accepted 04.06.20; published 21.09.20.

Please cite as:
Monteiro-Guerra F, Signorelli GR, Rivera-Romero O, Dorronzoro-Zubiete E, Caulfield B
Breast Cancer Survivors' Perspectives on Motivational and Personalization Strategies in Mobile App–Based Physical Activity Coaching Interventions: Qualitative Study
JMIR Mhealth Uhealth 2020;8(9):e18867
URL: https://mhealth.jmir.org/2020/9/e18867
doi:10.2196/18867
PMID:32955446

Original Paper

Development and Evaluation of an Accelerometer-Based Protocol for Measuring Physical Activity Levels in Cancer Survivors: Development and Usability Study

Tracy E Crane[1*], PhD, RDN; Meghan B Skiba[2*], MS, RDN; Austin Miller[3], PhD; David O Garcia[2], PhD, FACSM; Cynthia A Thomson[2], PhD, RDN

[1]Department of Biobehavioral Health Sciences, College of Nursing, University of Arizona, Tucson, AZ, United States

[2]Department of Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

[3]Department of Biostatistics, Roswell Park Cancer Institute, Buffalo, NY, United States

[*]these authors contributed equally

Corresponding Author:
Tracy E Crane, PhD, RDN
Department of Biobehavioral Health Sciences
College of Nursing
University of Arizona
1305 N Martin Ave
Tucson, AZ, 85721
United States
Phone: 1 5203310120
Email: tecrane@email.arizona.edu

Abstract

Background: The collection of self-reported physical activity using validated questionnaires has known bias and measurement error.

Objective: Accelerometry, an objective measure of daily activity, increases the rigor and accuracy of physical activity measurements. Here, we describe the methodology and related protocols for accelerometry data collection and quality assurance using the Actigraph GT9X accelerometer data collection in a convenience sample of ovarian cancer survivors enrolled in GOG/NRG 0225, a 24-month randomized controlled trial of diet and physical activity intervention versus attention control.

Methods: From July 2015 to December 2019, accelerometers were mailed on 1337 separate occasions to 580 study participants to wear at 4 time points (baseline, 6, 12, and 24 months) for 7 consecutive days. Study staff contacted participants via telephone to confirm their availability to wear the accelerometers and reviewed instructions and procedures regarding the return of the accelerometers and assisted with any technology concerns.

Results: We evaluated factors associated with wear compliance, including activity tracking, use of a mobile app, and demographic characteristics with chi-square tests and logistic regression. Compliant data, defined as ≥4 consecutive days with ≥10 hours daily wear time, exceeded 90% at all study time points. Activity tracking, but no other characteristics, was significantly associated with compliant data at all time points (P<.001). This implementation of data collection through accelerometry provided highly compliant and usable activity data in women who recently completed treatment for ovarian cancer.

Conclusions: The high compliance and data quality associated with this protocol suggest that it could be disseminated to support researchers who seek to collect robust objective activity data in cancer survivors residing in a wide geographic area.

(JMIR Mhealth Uhealth 2020;8(9):e18491) doi:10.2196/18491

KEYWORDS

wearable electronic devices; physical activity; cancer survivors; activity trackers; mobile phone

Introduction

Background

For cancer survivors, it has been demonstrated that physical activity has positive effects on psychosocial and physical outcomes, including weight management, quality of life, fatigue, emotional well-being, and sleep as well as social, cognitive, and physical functioning [1-3]. However, self-reported levels of physical activity among ovarian cancer survivors are low, with approximately 20% of women meeting the recommended 150 min of moderate-to-vigorous physical activity per week [4,5]. Measurement and assessment of physical activity in cancer survivors remain a challenge in randomized controlled trials, wherein self-reported questionnaires are most commonly used [6]. Subjective self-report of physical activity is wrought with significant recall and measurement error bias, as many study participants overreport moderate- and vigorous-intensity physical activities [7]. In previous trials, participants not only overreported physical activity but also there was a concurrent tendency to underreport sedentary time [8]. Furthermore, many self-report instruments fall short in capturing light-intensity physical activity [9], a major source of activity in adults, particularly cancer survivors [10]. In fact, in the Women's Health Study, fewer than 50% of women met physical activity guidelines as measured using accelerometry compared with approximately 67% women from self-reported physical activity [11]. Objective measurement of activity is, therefore, considered the gold standard for assessing physical activity exposure for all levels of intensity (rest, sedentary, light intensity, and moderate-to-vigorous intensity) in clinical trials and epidemiological studies.

Accelerometry provides the opportunity to objectively evaluate the minutes of activity per day, intensity, and total energy expenditure of physical activity as well as sedentary time, known factors associated with cancer risk [12,13]. Accelerometer protocols have been successfully implemented in population-level surveillance studies [10]. However, the existing information related to distance accelerometer methods (eg, mail based) has suggested poor compliance, with minimal improvement in recent years despite the growing use of this technique in assessing physical activity [14]. The use of accelerometry to capture objective activity data is particularly scarce among randomized and specifically lifestyle intervention trials in cancer survivors. A recent review highlighted that methods of accelerometry data collection among cancer survivors varied and were inconsistent, with limited studies reporting necessary details regarding data collection, compliance, and processing of data that would support replication of the research [15]. Robust measurement of physical activity using objective methods in cancer survivors is of high importance. Cancer survivors have unique needs as a result of cancer and subsequent treatment, including but not limited to fatigue, ostomies, abdominal pain, and chemotherapy-induced peripheral neuropathy [4,16,17], all of which can be a barrier to physical activity.

Objectives

To improve the rigor of physical activity assessment in both epidemiologic and intervention studies, a standardized protocol for collecting reliable and valid mail-based accelerometer data is warranted. In this paper, we describe a mail-based protocol for prospective collection of accelerometry data in a convenience sample of 580 ovarian cancer survivors enrolled in a diet and physical activity intervention. Of note, study participants resided in 48 US states, suggesting that this protocol may be applicable and implemented in studies that recruit participants across wider geographic areas. Finally, the protocol was tested across multiple study time points to demonstrate compliance over time, a critical aspect of longitudinal research.

Methods

Wearable Electronic Device

The Actigraph GT9X Link is a validated triaxial accelerometer [18] that includes a gyroscope, magnetometer, secondary accelerometer, and Bluetooth capability [19] manufactured by ActiGraph, LLC. The Actigraph GT9X Link uses the same validated algorithms from its predecessor, GT3X, which was validated through indirect calorimetry [20]. Compared with GT3X, GT9X Link captures different step counts [21] but still provides comparable data and estimates of activity intensity and sedentary time [22]. Few studies have implemented GT9X Link to assess physical activity to date, but none have been in cancer survivors [23-25].

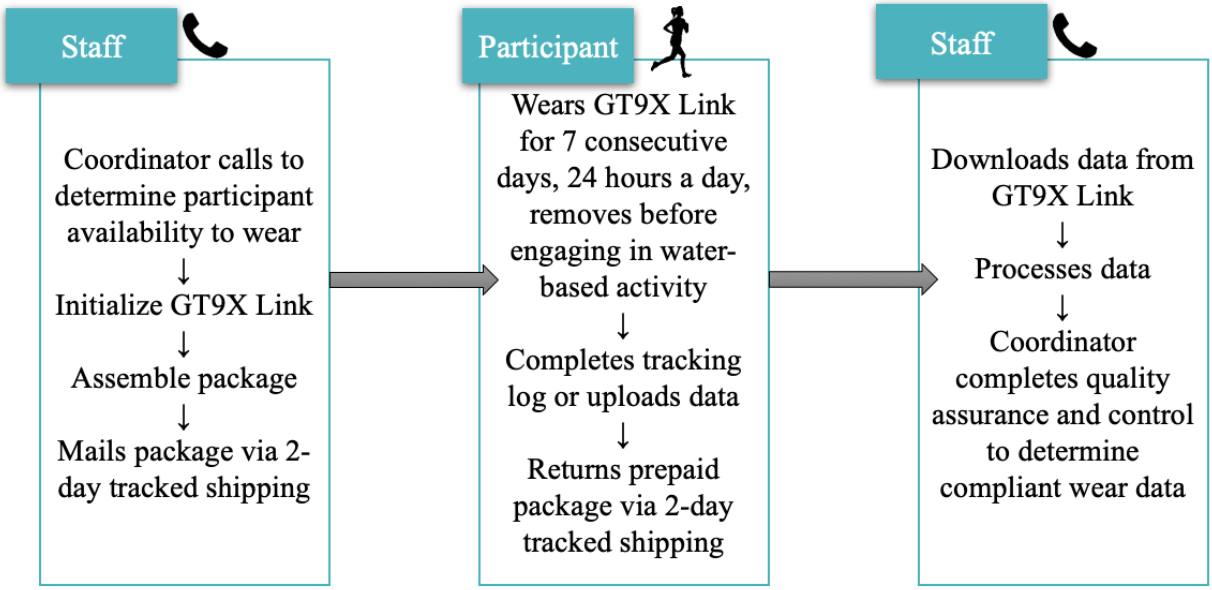
Physical activity was measured within the implementation of the GOG/NRG 0225 Lifestyle Intervention for oVarian cancer Enhanced Survival (LIvES) study (NCT00719303). Briefly, the LIvES study is a randomized controlled trial that tested a 24-month lifestyle intervention (high vegetable and fiber, low-fat diet with daily physical activity goals) compared with an attention control (general health education) on ovarian cancer progression-free survival. Eligible participants were in 6 weeks to 6 months postcancer treatment for stages II to IV disease and were randomized in a ratio of 1:1 to intervention versus control groups, stratified by receipt of consolidation therapy. Women were enrolled at NRG Oncology clinic sites nationwide. Intervention components are delivered via telephone by trained health coaches from the University of Arizona Cancer Center (UACC). Protocol and related methodologies as well as retention approaches for this study were previously published [26]. Measures included self-report physical activity assessment via the validated Arizona Physical Activity Questionnaire (APAQ) [27] and repeat accelerometer-measured activity at time points aligned with the self-report measure (baseline, 6 months, 12 months, and 24 months) as a measure of intervention adherence. The APAQ also included self-report of participants' height and weight, which was used to calculate their BMI. Participant demographics were collected using a standardized form at enrollment.

Accelerometer Protocol

The accelerometer protocol was developed using manufacturer instructions, a review of published literature [28-32], and prior experience of study investigators. The protocol development

process included the creation of standardized instructional materials, a staff training program, implementation strategies, and data quality assessments. The objective activity was assessed using the GT9X Link accelerometer at 4 study time points: baseline, 6 months, 12 months, and 24 months (Figure 1). Participants were asked to wear GT9X Link on their nondominant hip for a continuous period of 7 days, 24 hours a day, with the exception of water-based activities such as showering or swimming.

Figure 1. Participant communication schema for accelerometry data capture in the Lifestyle Intervention for oVarian cancer Enhanced Survival study.



Participant Instructional Materials

The instructional materials for this protocol included a cover letter that summarized the procedures, guidance on wear, troubleshooting, and tracking logs for wear time. Before implementing the protocol, study staff and 2 individuals outside of the study (who were of the same age as study participants) beta-tested the materials to determine usability and acceptability of the instructional packet, tracking log, and mobile device app. Feedback from usability and acceptability testing was incorporated into the final study documents. The instruction packet was revised, as technology related to GT9X Link (such as chargers and waist clips) was updated in the inventory by ActiGraph.

An introductory letter was included in each package that oriented participants to the activity monitor portion of the study and acknowledged and thanked them for their study participation to date, specifically in relation to data collection at prior time points. The letter detailed the list of materials included in the mailed package and provided a study telephone number to call if they had any questions. The instructional booklet served as a resource for participants and included written instructions with photos that described how to (1) charge the activity monitor, (2) wear the activity monitor on the nondominant hip, (3) complete the tracking log, (4) download the CentrePoint Study Admin Sync app on a mobile device for data upload, and (5) prepare the package for return to the data collection center at the end of the 7-day wear period. At baseline, for participants who provided an email address, an instructional video that detailed the same information contained in the written instructional booklet was sent via email with the embedded web link. A monthly calendar was provided as a resource for participants to mark which days of the week the activity monitor was required to be worn and the date to return the accelerometer. Finally, a checklist of related Actigraph GT9X Link wear procedures was included as a tool to remind the participants of each step in the process and to promote inclusion of the activity log in the return package to the UACC Behavioral Measurement and Interventions Shared Resource (BMISR).

Self-Monitoring

A tracking log was provided to all participants to self-monitor wear time and record all accelerometer wear time on a daily basis for the 7-day wear period. The tracking log detailed self-reported measures of wake and sleep times as well as removal times of the accelerometer and indicated if the data were uploaded via the CentrePoint Study Admin Sync app. These data were used as quality assurance measures to determine the correct wear times for individual participants (which were then verified by accelerometer readings) as well as compliance with self-monitoring behaviors. Engagement with the CentrePoint mobile app, defined as a mobile upload of data during the 7-day wear period at least once, was confirmed by the researcher through the backend of the CentrePoint mobile app.

Staff Training

Study staff working with accelerometer data collection were trained and adhered to a standard operating procedure (SOP), specifically created for objective physical activity data collection in the context of the study. The SOP included guidance for staff in relation to initializing the accelerometer and downloading data from the Actigraph GT9X Link as well as the procedure for contacting participants (eg, phone script and number of call attempts) and mailing the packages. These SOPs were informed

by previous studies that used the Actigraph GT9X Link accelerometers for the collection of physical activity data and further adapted for the study population in an effort to enhance compliance beyond prior reports [28-32].

Mailing Approaches

After confirming the availability to wear the accelerometer, the study staff created packages to send to the participant's home address. The packages included (1) introduction letter, (2) tracking log, (3) instruction packet, (4) charging pack, (5) return checklist, (6) monthly study activity calendar, and (7) return postage paid preaddressed return envelope. The United States Postal Service (USPS) Priority Flat Rate padded envelopes were used for sending and receiving packages. At the time of this study, USPS Priority mail provided 2-day shipping to the Continental United States, which was chosen to reduce the likelihood of an accelerometer losing charge in transit and expedite the wearing process to ensure that data could be collected in an appropriate time frame. USPS tracking numbers were assigned to each outgoing and incoming package and documented. Study staff would access updates of the location of the package through a web-based tool maintained by the USPS. The expected battery life of GT9X Link is 10 days; however, to ensure that battery life lasted through the duration of the shipping and wear period, a charging pack was included with the accelerometer. The window to open the accelerometer package and initiate wear time was ±2 weeks (14 days) at each time point. An extended window of an additional 4 weeks was used in extenuating circumstances such as extended travel, illness, or reissue of the accelerometer because of damage or loss. At the end of the wear duration, participants mailed the activity monitor, charging pack, and tracking log in the prepaid and addressed return envelope provided in the original package.

Communication

At each study time point, participants were contacted by trained study staff from UACC-BMISR via telephone. During this call, the participant was queried as to when and whether they would be available to wear the accelerometer for 7 consecutive days. If they were agreeable to wearing the accelerometer through verbal consent and confirmation, a follow-up call occurred within 2 days of confirmed delivery of the package to determine receipt by the participant and provide a detailed explanation of wear and documentation procedures. Predetermined wear dates were estimated by the study staff, including setting a return date estimated and documented as the last day of anticipated wear. An additional 4 days were added for return shipping. Participants were instructed to wear the accelerometer continuously for 24 hours a day for a total duration of 7 days, including during sleep, with the exception of when bathing, showering, or swimming. Participants were instructed to first fully charge the activity monitor upon receipt before wearing (and if the battery life dropped to <10%) and to start the wear time upon awakening the next day and continue to wear through wake time after day 7. For the duration of wear time, participants were requested to complete a tracking log that included date, time awake, time asleep, times they took off the device, and the reason the device was removed. If 1 of the wear days was missed, they were asked to wear the device for an additional day and document on the tracking log reason for nonwear (eg, clinic appointment, illness, and forgot).

A few areas of protocol implementation required troubleshooting to enhance compliance. The first was to reach the participant for the initial accelerometer distribution. Following the SOP, the study staff attempted to contact a participant twice a week for 2 consecutive weeks, after which the study coordinator would contact the clinic to ensure that contact information was correct before continuing further attempts to contact the participant. Second, select participants who delayed in returning the accelerometer. If an accelerometer was not received within 7 days of the anticipated return date, the participant would be called by the study coordinator and, if required, a voicemail was left. The expected return date was then updated. If still it was not received within 7 days of this extended return date, the participant was contacted again by telephone, and an email reminder was sent. If still not returned by 3 weeks after the initial expected return date, the oncology clinic where the participant gave consent for trial participation was contacted to confirm whether the participant was still active in the study and request that the clinic discusses the return of the accelerometer with the study participant during their next scheduled clinic visit. In addition, the study staff would attempt to contact the participant to return the accelerometer once a week for the next 4 weeks. If the accelerometer was not successfully returned after 2 consecutive months of contact attempts, the participant was considered protocol noncompliant, and a USPS-certified letter from the study's principal investigator was mailed to the participant's residential address to request the return of materials and provided instructions on how to do so.

Data Collection and Quality Assessment

Data Collection and Capture

The study staff tracked contact attempts with participants at each time point using an encrypted shared spreadsheet. This spreadsheet contained no personal identifying information, only the study participant ID numbers. Name, contact number, email, and mailing address were stored separately on the Health Insurance Portability and Accountability Act (HIPAA)–secure study platform and database [33]. The enrollment date of the participant was added to this spreadsheet and used to calculate the open windows for subsequent time points. It also included the initialization date of the accelerometer, mailing date, and tracking numbers of the packages as well as return date, download date, and whether or not compliant data were collected. Any special notes about the participant during the specified time point were documented on the spreadsheet. Of note, if a participant refused to wear or was unable to wear an accelerometer at one time point, this was documented, and they were still queried at subsequent time points if they remained active in the study. Refusal to wear was defined as the participant declining to wear the accelerometer for any reason (eg, not having available time, inconvenience, conflicts with religious holidays, or discomfort with wearing the accelerometer). Unable to wear was defined as participants expressing willingness to wear an accelerometer but not having physical capacity to wear during open windows (eg, surgery, hospitalization, illness or injury, or natural disaster). Participants

were removed from being contacted at future study time points if the participant had reached a study endpoint (eg, disease progression, other diseases, lost to follow-up, or withdraw of participation) or was previously protocol noncompliant (eg, did not return the accelerometer after 2 consecutive months of contact attempts).

Data collection included the use of the Actigraph GT9X Link Study Admin Sync, a web-based study management platform for accelerometers. This platform permits participant connection through CentrePoint Study Admin Sync, a mobile app available on iOS and Android devices that gives participants the option to upload their accelerometer data in real time to a secure HIPAA–compliant cloud-based study database [34]. At the time of developing this protocol, there was no literature available regarding participant uptake of the use of this mobile app. Therefore, the integration of this platform was an opportunity to explore the use of a mobile app as a data collection tool to inform future research. Participants were asked whether they used a mobile device (eg, smartphone or tablet) and whether they would be willing to download the mobile app that connects the mobile device to the accelerometer through Bluetooth connectivity and upload the data through user response to the associated study database in the Actigraph GT9X Link Study Admin Sync. During the follow-up call, if the participant opted to use the mobile app, instructions for downloading and installing the app on their mobile device were covered in detail. To connect Actigraph GT9X Link to the app, a unique 5-digit code was given to the participant via telephone. Participants entered this code on their device within a 2-min time frame before the code expired. This mobile app was optional for the participant to upload their data at the end of each waking period.

In an attempt to rigorously collect objective physical activity data, provisions were made to reissue an accelerometer if the accelerometer was lost and data could be captured within the predesignated extended wear time frame window of 4 weeks. The same protocol for initializing and mailing the replacement accelerometer was followed.

All accelerometers and accessories were sanitized by the study staff upon return to UACC-BMISR. A separate spreadsheet was used by the study staff to track all accelerometers in a study-specific inventory. If a participant reported a problem with their accelerometer or charger, this was documented by the study staff and investigated upon return of the accelerometer. If the problem could not be resolved, the accelerometer or charging dock would be removed from the mailing rotation and replaced.

The accelerometers were initialized, and the data were downloaded using the Actigraph GT9X Link CentrePoint Study Admin System. Weight and age were entered for each participant during initialization, and the time zone of the participant was documented by the study staff. Participants were blinded to the physical activity feedback from the device; the screen was programmed to only display battery life, date, and time.

Data Quality Assessment

As part of quality control, the study staff visually inspected all processed data for compliance. Compliance with the accelerometer wear time was defined by the investigator group before study initiation. These cut-off points have been previously validated in adults; no specific cut-off points currently exist for cancer survivors [35]. Specifically, wear compliance was set at ≥4 days, with ≥10 hours of daily wear time [36,37]. Initial data from the accelerometer were downloaded and exported to ActiLife (version 6.12) software. Accelerometer data were processed using 60-second epochs and Freedson [38] cut-off points.

Statistical Analysis

Descriptive data are reported as frequency, means, or medians. Potential predictors of compliant data, including demographic characteristics and self-monitoring behavior, at baseline were evaluated using logistic regression models. As the sample size decreased longitudinally from participants reaching a predetermined study endpoint, logistic regression models were not repeated for subsequent time points. Comparison of factors related to self-monitoring behaviors and compliant accelerometry data collected at each time point were conducted using the Pearson chi-square tests. All statistics were completed in Stata 16 (StataCorp LLC).

Results

Overview

From July 2015 to December 2019, 580 cancer survivors active in the study were contacted by the study staff at enrollment to initiate the accelerometer procedures, 12.9% (75/580) provided an email address, and the majority preferred to communicate via telephone. Over the 4 time points, a total of 1337 individual accelerometer mailings were completed. At baseline, 98.4% (571/580) women were available and willing to wear the accelerometer—95.2% (533/560) study participants were compliant with the wear protocol and had usable data, and 5% (27/560) had insufficient wear time. For the 3.5% (20/580) participants without the accelerometer wear time at baseline, reasons for nonparticipation included participants who refused to wear the accelerometer (n=3), unable to wear because of illness or natural disasters (n=8), and off-study before completing baseline assessments (n=9).

At subsequent time points, participants who demonstrated protocol noncompliance (eg, not returning the accelerometer at a previous time point) were not asked to collect follow-up accelerometer data. In addition, if a participant reached a study endpoint, no further measurements of activity were completed. At 6 months, 95.2% (394/414) of individual participants were sent accelerometers, resulting in the acquisition of compliant data from 90.9% (358/394) of the active sample. At 12 months, 260 active participants were sent accelerometers for repeat activity measurement, with 95.4% (248/260) of women providing compliant data. At 24 months, 123 eligible participants were sent accelerometers, resulting in 96.7% (119/123) of the active sample that provided compliant data.

The details of the accelerometer data compliance by time points are outlined in Table 1.

Across all time points, 49 accelerometers required a reissue for participant data collection. The majority (32/49, 65%) of the reissued accelerometers were at baseline. Of these reissues, 25% (12/49) were because of depleted battery life upon arrival, 20% (10/49) for software malfunction, and 20% (10/49) were lost by either the participant or in transit. Removal from mailing rotation and replacement of inventory for accelerometers that encountered issues and nonfunctional charging docks were 16 and 67, respectively.

Participants engaged in the accelerometer protocol were representative of the overall Lifestyle Intervention for oVarian cancer Enhanced Survival study population in terms of demographic and clinical characteristics [26]. On average, participants were aged 60.1 (SD 9.3) years, and the majority were non-Hispanic college graduates with a normal BMI (Table 2). The results from logistic regression models of age, education, ethnicity, and BMI as well as self-monitoring behaviors indicated that tracking log completion was the greatest predictor of compliant data at baseline (Table 3). Odds ratios (ORs) were significantly higher (OR 54.71, 95% CI 17.05-175.59; $P<.001$) for having compliant data if a participant completed a corresponding tracking log compared with those who did not complete the tracking log during the 7-day wear period. Other factors, including age, education, ethnicity, BMI, and mobile app engagement, were not significantly associated with compliant data at baseline.

Table 1. Details of the accelerometer wear time compliance in 580 participants by time points on the Lifestyle Intervention for oVarian cancer Enhanced Survival accelerometer protocol.

Study sample[a]	Baseline[a], n (%); N	6 months[a], n (%); N	12 months[a], n (%); N	24, months[a], n (%); N
Total active[b] sample	571 (98.4); 580	414 (71.4); 580	284 (49.0); 580	154 (28.7); 580
Agree to wear[c]	560 (98.1); 571	394 (95.2); 414	260 (91.5); 284	123 (79.9); 154
Compliant data[d]	533 (95.2); 560	358 (90.9); 394	248 (95.4); 260	119 (96.7); 123
Noncompliant data[d]	27 (5); 560	36 (9); 394	12 (5); 260	4 (3); 123
Refused to wear[e]	3 (<1); 571	16 (4); 414	21 (7); 284	24 (16); 154
Unable to wear[f]	8 (1); 571	4 (1); 414	3 (1); 284	5 (3); 154
Total inactive[g]	9 (2); 580	166 (28.6); 580	296 (51.0); 580	384 (71.6); 536
Protocol noncompliant[h]	N/A[i]	14 (8); 166	25 (8); 296	27 (7); 384
Study endpoint[j]	9 (100); 9	152 (91.6); 166	271 (91.6); 296	357 (93.0); 384

[a]Values may not add up to 100% because of rounding.

[b]Active participants included those still on study time point and eligible to wear an activity monitor at each time point. At 24 months, 44 women were not yet at the study time point and, therefore, were not included in the sample size.

[c]Agree to wear is defined as a participant who provided verbal consent and confirmation of availability to wear the accelerometer via phone.

[d]Compliant data were defined as ≥4 consecutive days, with ≥10 hours of daily wear time. Percent is calculated by the number of all participants who provided verbal consent and were available to wear the accelerometer. Participants who were not active in the study were not included in the denominator for data compliance.

[e]Refusal to wear is defined as a participant who declined to wear the accelerometer for any reason.

[f]Unable to wear is defined as a participant who expressed willingness to wear the accelerometer but did not have the physical capacity to wear during the open window.

[g]Inactive participants included those who were protocol noncompliant or had reached a study endpoint.

[h]Protocol noncompliant is defined as a participant who did not return the accelerometer after 2 consecutive months of contact attempts.

[i]N/A: not applicable.

[j]Participants who reached a study endpoint, defined as disease progression, other diseases, lost to follow-up, or withdraw of participation, were not asked to wear the accelerometer at any future time points.

Table 2. Baseline characteristics for women who were enrolled in the Lifestyle Intervention for oVarian cancer Enhanced Survival accelerometer protocol at baseline (n=580).

Characteristics	Values
Age at enrollment (years), mean (SD)	60.07 (9.3)
Education[a], n (%)	
High school or less	75 (14.0)
Some college education	147 (27.4)
College graduate	314 (58.6)
Ethnicity[a], n (%)	
Hispanic	34 (6.4)
Non-Hispanic	500 (93.6)
BMI class (kg/m^2), n (%)	
Normal (18.5-24.9)	211 (36.4)
Overweight (25.0-29.9)	203 (35.0)
Obese (≥30.0)	166 (28.6)

[a]Missing data <8%.

Table 3. Predictors of baseline compliant data and association with demographic and anthropometric characteristics (n=522).

Characteristics[a,b]	Odds ratio (95% CI)	*P* value
Age	1.00 (0.95-1.05)	>.99
Education		
High school or less	1.0[c]	N/A[d]
Some college education	3.27 (0.74-14.44)	.12
College graduate	2.60 (0.68-9.92)	.16
Ethnicity		
Non-Hispanic	1.0	N/A
Hispanic	0.89 (0.15-5.25)	.90
BMI class		
Normal	1.0	N/A
Overweight	1.17 (0.33-4.13)	.80
Obese	0.88 (0.25-3.07)	.84
Tracking log completion	54.71 (17.05-175.59)	<.001
Mobile app engagement	2.79 (0.56-13.84)	.21

[a]Using data available for participants who agreed to wear an accelerometer at baseline.

[b]Only participants with nonmissing values for all variables are included in the logistic regression model.

[c]These are all provided with the exception of the referent group, which would *not* have a CI—this is the group we are comparing against to determine the OR and CI.

[d]N/A: not applicable.

Compliance With Tracking Logs and Mobile App Engagement (Self-Monitoring)

At baseline, 88.6% (496/560) of the tracking logs were returned. At subsequent time points, 83.0% (327/394) were returned at 6 months, 84.2% (219/260) were returned at 12 months, and 86.2% (106/123) were returned at 24 months. The most common reason reported for the removal of the Actigraph GT9X-Link on the tracking logs was shower or bath, aligned with the wear protocol. All other reasons for removal accounted for <10% of the reported removal for all days and included allowing the monitor to charge, swimming, or medical examination. Completion of the tracking log was significantly associated with compliant accelerometer wear data at all time points (*P*<.001, Table 4).

Table 4. Prevalence of tracking log completion and manufacturer mobile app engagement with compliant accelerometry data collection among the active Lifestyle Intervention for oVarian cancer Enhanced Survival participants who agreed to wear an accelerometer at each study time point.

Time point[a,b]	Value, n (%)	*P* value
Baseline (n=560)		
Tracking log	496 (88.6)	<.001
Mobile app	198 (35.4)	.002
Both	187 (33.4)	<.001
6 months (n=394)		
Tracking log	327 (83.0)	<.001
Mobile app	78 (19.8)	.06
Both	60 (15.2)	.03
12 months, n=260		
Tracking log	219 (84.2)	<.001
Mobile app	41 (15.8)	.53
Both	35 (13.5)	.16
24 months (n=123)		
Tracking log	106 (86.2)	<.001
Mobile app	13 (10.6)	.49
Both	12 (9.8)	.50

[a]Percent values are calculated by dividing by the number of active participants who agreed to wear an accelerometer at each time point.

[b]Analyses were conducted using the Pearson chi-square tests.

At baseline, 35.4% (198/560) of the participants opted to use the mobile app; the average frequency of uploads during the 7-day wear period was 5 (median 6). At 6, 12, and 24 months, 19% (78/123), 15% (41/123), and 11% (13/123) of the participants repeated the use of the app, respectively. The average frequency of uploads at 6, 12, and 24 months was 5 (median 6), 6 (median 6), and 6 (median 6), respectively. The use of the mobile app was significantly associated with compliant data at baseline only (P=.002). Completion of both the tracking log and mobile app upload was significantly associated with compliant accelerometry data collected at baseline and 6 months (P=.001 and P=.03, respectively). Fewer participants completed both the tracking log and uploaded data through the mobile app at the following 12- and 24-month time points, with the majority of participants choosing to complete the tracking log only.

Discussion

Principal Findings

Given the current interest in objective measurement of physical activity levels in cancer survivors and reports of over- and under-reporting of physical activity using validated self-reported measures, a rigorous protocol, which includes SOPs, to successfully send, receive, collect, and download accelerometer data is recommended. A 2018 review of accelerometer-based activity monitoring in cancer survivors suggested that current efforts lack standardization in relation to the methodology used and further details of accelerometer data collection methods are lacking, making replication at best challenging [15]. To address this gap in current evidence, this study provides an adaptable, detailed protocol for consideration across trials, particularly those that are distance-delivered. On the basis of our findings, these protocols should include frequent communication via telephone or email, multiple methods of instruction for how to use the accelerometer (eg, video, paper, and telephone-based education), and a detailed mail tracking system. To increase the likelihood of receiving compliant and usable data, protocols should include having the participants upload their data in real time or record wear time on a standardized tracking form. Here, we report the study protocol, implementation, and initial compliant data for using accelerometers in a large, multisite national lifestyle intervention in a trial of ovarian cancer survivors. Data were collected from participants in 48 US states for objective measurement of activity. The LIvES research team has developed a reproducible protocol that can be used by others for the implementation of accelerometers in future research trials.

Comparison With Prior Works

The majority (98%) of the data were collected within the predesignated windows for data collection. Agreement to wear the accelerometer did decrease over time among all protocol active participants. Our goal was to collect data from all active women within the subsample at all time points. However, a small percentage of the women were protocol noncompliant at previous time points or refused or were unable to wear the accelerometer (Table 1). Of those who provided verbal consent and availability to wear the accelerometer, 95.2% (533/560), 90.9% (358/394), 95.4% (248/260), and 97.6% provided compliant data at baseline, 6, 12, and 24 months, respectively. Our compliance rates for accelerometry were similar to

large-scale prospective cohort studies in older women with single time point measurements [36] and were above the 71.6% wear time compliance estimates from adults wearing thigh and back-placed accelerometers [39] and much higher than estimates of 62.6% among children who participated in the National Health and Nutrition Examination Survey [40]. Compliant data on distance-delivered methodologies are sparse.

At the time of the development of this protocol, the GT9X was validated only for hip-worn measurements. Although the GT9X has since been validated for wristwear [18], it is unknown how this may have affected compliance in our sample. Among a large sample of European adults with a single time point of collection, 93.3% had valid data for wrist-worn accelerometers [41]. Comparatively, we demonstrated 95.2% valid compliant data from hip-worn accelerometers in our sample at baseline. The current literature suggests a poor correlation between wrist and hip-worn accelerometer counts per min; therefore, comparison of estimates from wear at the 2 different sites should be interpreted with caution [42,43]. To maintain consistency in the data collected, all participants were asked to wear GT9X on their nondominant hip at all time points. Furthermore, the hip placement of GT9X demonstrates better step count accuracy [18], which is an *a priori* behavior outcome for the trial. In addition, the quality of the data generated using the protocol will allow for time in sedentary bouts, light activity, and the development of population-specific cut-off points that accommodate the described reduction in peak oxygen consumption and relative intensity of activities among cancer survivors [44,45].

Contrary to other studies that may suggest that obese women have higher nonwear time of accelerometers [46], our results indicate that BMI did not have a significant effect on compliant data in ovarian cancer survivors. Tracking log completion remained strongly associated with compliant wear data across all time points. Calls at 6, 12, and 24 months became more streamlined and shorter in length, as many participants were already familiar with the protocol and felt comfortable with the device without additional telephone support at these time points. High compliance overall may reflect the older female sample of cancer survivors motivated to participate in a 24-month lifestyle intervention.

The use of the mobile app to upload data was optional, and 35.4% of women opted to use this app. We noted that age, though not statistically significant, influenced the use of the mobile app, suggesting that this approach may have higher adherence for younger cancer survivors. The mobile app allowed participants to upload data but they could not see their activity data to keep data collection blinded. This may have influenced the motivation for repeated app engagement. Over time, participants engaged less with the mobile app, results that are similar to the patterns previously reported in the literature for other health-related smartphone apps [47]. This warrants further studies, including new strategies to promote the continued use of technology. Of note, our instruction call notes from participants indicated that many participants did not find any benefit or did not feel like there was time available to complete both the tracking log and mobile app. An additional consideration in applying the findings of this research is our focus on a convenience sample of ovarian cancer survivors. Although we anticipate that the protocol will perform similarly in other cancer survivor populations, this is not a standard, particularly in relation to diversity in sex and education as well as cancer-related symptom burden and comorbidities.

Future Directions

Beyond the scope of this protocol, but important to future research in this area, population-relevant cut-off points and algorithms need to be established for cancer survivors for both wrist and hip-worn accelerometers. The field of accelerometry is rapidly emerging, especially regarding standardized cut-off points. For data analysis related to objective physical activity data captured in this protocol, standard cut-off points as well as vector magnitude and total activity counts will be evaluated by the research team. The current literature suggests that vector magnitude may better discriminate between sedentary and light physical activity in women aged older than 60 years [48], although among breast cancer survivors, total activity counts may provide better estimates of moderate-to-vigorous physical activity [49]. These population-relevant cut-off points will allow for a more accurate interpretation of accelerometer data for the older female cancer survivor population; however, information remains limited for cut-off points specific to ovarian cancer survivors. Importantly, both age and the presence of comorbidities can influence accelerometer cut-off points [48], confounders that are highly relevant to this population.

In summary, we have developed a detailed protocol and related materials for collecting the accelerometry data from a large sample of cancer survivors who reside across the United States. This protocol has resulted in the acquisition of a robust data set for future analysis of physical activity in this population. This protocol and the related materials that were issued to participants are available through UACC-BMISR consultation services [50] in support of future research studies designed to capture repeated measures of activity in this vulnerable population.

Acknowledgments

The authors wish to thank all the work and efforts of the staff and the undergraduate and graduate student research team who worked on this project from its inception: Natasha Burrell, Kelly Carroll, Sarah Lavelle, Sophia Perkin, Jennifer Shipley, McKenna Scheldt, Madeline Schneider, Patrice Tillman, Amanda Younger, and Samantha Werts. We also thank the 580 ovarian cancer survivors who dedicated their time and provided data for this research. This work was supported by the National Cancer Institute grants to NRG Oncology (U10 CA 180822), NRG operations (U10 CA180868), and UG1CA189867 (NCORP). Additional funding for this research was provided by the NRG Grant (R305B160003), the National Institutes of Health National Cancer

Institute (1R01 CA186700), the BMISR at the UACC Support Grant (P30 CA023074), and the West Valley Ovarian Cancer Alliance.

Conflicts of Interest

None declared.

References

1. Mishra SI, Scherer RW, Geigle PM, Berlanstein DR, Topaloglu O, Gotay CC, et al. Exercise interventions on health-related quality of life for cancer survivors. Cochrane Database Syst Rev 2012 Aug 15(8):CD007566 [FREE Full text] [doi: 10.1002/14651858.CD007566.pub2] [Medline: 22895961]
2. Garcia DO, Thomson CA. Physical activity and cancer survivorship. Nutr Clin Pract 2014 Dec;29(6):768-779 [FREE Full text] [doi: 10.1177/0884533614551969] [Medline: 25335787]
3. Smits A, Smits E, Lopes A, Das N, Hughes G, Talaat A, et al. Body mass index, physical activity and quality of life of ovarian cancer survivors: time to get moving? Gynecol Oncol 2015 Oct;139(1):148-154. [doi: 10.1016/j.ygyno.2015.08.005] [Medline: 26291650]
4. Mizrahi D, Naumann F, Broderick C, Samara J, Ryan M, Friedlander M. Quantifying physical activity and the associated barriers for women with ovarian cancer. Int J Gynecol Cancer 2015 May;25(4):577-583. [doi: 10.1097/IGC.0000000000000349] [Medline: 25628105]
5. Kushi LH, Doyle C, McCullough M, Rock CL, Demark-Wahnefried W, Bandera EV, American Cancer Society 2010 Nutrition and Physical Activity Guidelines Advisory Committee. American cancer society guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin 2012;62(1):30-67 [FREE Full text] [doi: 10.3322/caac.20140] [Medline: 22237782]
6. Irwin ML, Ainsworth BE. Physical activity interventions following cancer diagnosis: methodologic challenges to delivery and assessment. Cancer Invest 2004;22(1):30-50. [doi: 10.1081/cnv-120027579] [Medline: 15069762]
7. Vassbakk-Brovold K, Kersten C, Fegran L, Mjåland O, Mjåland S, Seiler S, et al. Cancer patients participating in a lifestyle intervention during chemotherapy greatly over-report their physical activity level: a validation study. BMC Sports Sci Med Rehabil 2016;8:10 [FREE Full text] [doi: 10.1186/s13102-016-0035-z] [Medline: 27099757]
8. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc 2014 Jan;46(1):99-106. [doi: 10.1249/MSS.0b013e3182a0595f] [Medline: 23793232]
9. Dowd KP, Szeklicki R, Minetto MA, Murphy MH, Polito A, Ghigo E, et al. A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. Int J Behav Nutr Phys Act 2018 Feb 8;15(1):15 [FREE Full text] [doi: 10.1186/s12966-017-0636-2] [Medline: 29422051]
10. Thraen-Borowski KM, Gennuso KP, Cadmus-Bertram L. Accelerometer-derived physical activity and sedentary time by cancer type in the United States. PLoS One 2017;12(8):e0182554 [FREE Full text] [doi: 10.1371/journal.pone.0182554] [Medline: 28806753]
11. Shiroma EJ, Cook NR, Manson JE, Buring JE, Rimm EB, Lee I. Comparison of self-reported and accelerometer-assessed physical activity in older women. PLoS One 2015;10(12):e0145950 [FREE Full text] [doi: 10.1371/journal.pone.0145950] [Medline: 26713857]
12. Zhong S, Chen L, Lv M, Ma T, Zhang X, Zhao J. Nonoccupational physical activity and risk of ovarian cancer: a meta-analysis. Tumour Biol 2014 Nov;35(11):11065-11073. [doi: 10.1007/s13277-014-2385-z] [Medline: 25095977]
13. Hildebrand JS, Gapstur SM, Gaudet MM, Campbell PT, Patel AV. Moderate-to-vigorous physical activity and leisure-time sitting in relation to ovarian cancer risk in a large prospective US cohort. Cancer Causes Control 2015 Nov;26(11):1691-1697. [doi: 10.1007/s10552-015-0656-7] [Medline: 26335264]
14. Montoye AH, Moore RW, Bowles HR, Korycinski R, Pfeiffer KA. Reporting accelerometer methods in physical activity intervention studies: a systematic review and recommendations for authors. Br J Sports Med 2018 Dec;52(23):1507-1516. [doi: 10.1136/bjsports-2015-095947] [Medline: 27539504]
15. Peddle-McIntyre CJ, Cavalheri V, Boyle T, McVeigh JA, Jeffery E, Lynch BM, et al. A review of accelerometer-based activity monitoring in cancer survivorship research. Med Sci Sports Exerc 2018 Sep;50(9):1790-1801. [doi: 10.1249/MSS.0000000000001644] [Medline: 29683922]
16. Stavraka C, Ford A, Ghaem-Maghami S, Crook T, Agarwal R, Gabra H, et al. A study of symptoms described by ovarian cancer survivors. Gynecol Oncol 2012 Apr;125(1):59-64. [doi: 10.1016/j.ygyno.2011.12.421] [Medline: 22155797]
17. Zhou Y, Irwin M, Ferrucci L, McCorkle R, Ercolano E, Li F, et al. Health-related quality of life in ovarian cancer survivors: results from the American cancer society's study of cancer survivors - I. Gynecol Oncol 2016 Jun;141(3):543-549 [FREE Full text] [doi: 10.1016/j.ygyno.2016.04.006] [Medline: 27072805]
18. Hwang J, Fernandez AM, Lu AS. Application and validation of activity monitors' epoch lengths and placement sites for physical activity assessment in exergaming. J Clin Med 2018 Sep 11;7(9):- [FREE Full text] [doi: 10.3390/jcm7090268] [Medline: 30208567]

19. ActiGraph Link GT9X. ActiGraph. 2018. URL: https://www.actigraph.nl/en/product/72/actigraph-link-gt9x.html [accessed 2019-08-03]
20. van Remoortel H, Raste Y, Louvaris Z, Giavedoni S, Burtin C, Langer D, PROactive Consortium. Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. PLoS One 2012;7(6):e39198 [FREE Full text] [doi: 10.1371/journal.pone.0039198] [Medline: 22745715]
21. John D, Morton A, Arguello D, Lyden K, Bassett D. 'What is a step?' differences in how a step is detected among three popular activity monitors that have impacted physical activity research. Sensors (Basel) 2018 Apr 15;18(4):- [FREE Full text] [doi: 10.3390/s18041206] [Medline: 29662048]
22. Montoye AH, Nelson MB, Bock JM, Imboden MT, Kaminsky LA, Mackintosh KA, et al. Raw and count data comparability of hip-worn actigraph GT3X+ and link accelerometers. Med Sci Sports Exerc 2018 May;50(5):1103-1112. [doi: 10.1249/MSS.0000000000001534] [Medline: 29283934]
23. Jaarsma T, Klompstra L, Ben Gal T, Boyne J, Vellone E, Bäck M, et al. Increasing exercise capacity and quality of life of patients with heart failure through Wii gaming: the rationale, design and methodology of the HF-Wii study; a multicentre randomized controlled trial. Eur J Heart Fail 2015 Jul;17(7):743-748 [FREE Full text] [doi: 10.1002/ejhf.305] [Medline: 26139585]
24. Noonan RJ, Fairclough SJ, Knowles ZR, Boddy LM. Context matters! sources of variability in weekend physical activity among families: a repeated measures study. BMC Public Health 2017 Apr 18;17(1):330 [FREE Full text] [doi: 10.1186/s12889-017-4232-9] [Medline: 28420363]
25. Coker-Bolt P, Downey RJ, Connolly J, Hoover R, Shelton D, Seo NJ. Exploring the feasibility and use of accelerometers before, during, and after a camp-based CIMT program for children with cerebral palsy. J Pediatr Rehabil Med 2017;10(1):27-36. [doi: 10.3233/PRM-170408] [Medline: 28339408]
26. Thomson CA, Crane TE, Miller A, Garcia DO, Basen-Engquist K, Alberts DS. A randomized trial of diet and physical activity in women treated for stage II-IV ovarian cancer: rationale and design of the lifestyle intervention for ovarian cancer enhanced survival (LIVES): an NRG oncology/gynecologic oncology group (GOG-225) study. Contemp Clin Trials 2016 Jul;49:181-189 [FREE Full text] [doi: 10.1016/j.cct.2016.07.005] [Medline: 27394382]
27. Staten LK, Taren DL, Howell WH, Tobar M, Poehlman ET, Hill A, et al. Validation of the Arizona activity frequency questionnaire using doubly labeled water. Med Sci Sports Exerc 2001 Nov;33(11):1959-1967. [doi: 10.1097/00005768-200111000-00024] [Medline: 11689750]
28. Lee I, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. Br J Sports Med 2014 Mar;48(3):197-201 [FREE Full text] [doi: 10.1136/bjsports-2013-093154] [Medline: 24297837]
29. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 2014 Jul;48(13):1019-1023 [FREE Full text] [doi: 10.1136/bjsports-2014-093546] [Medline: 24782483]
30. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. Br J Sports Med 2014 Mar;48(3):278-282. [doi: 10.1136/bjsports-2012-091410] [Medline: 22936409]
31. Broderick JM, Ryan J, O'Donnell DM, Hussey J. A guide to assessing physical activity using accelerometry in cancer patients. Support Care Cancer 2014 Apr;22(4):1121-1130. [doi: 10.1007/s00520-013-2102-2] [Medline: 24389829]
32. LaCroix AZ, Rillamas-Sun E, Buchner D, Evenson KR, Di C, Lee I, et al. The objective physical activity and cardiovascular disease health in older women (OPACH) study. BMC Public Health 2017 Feb 14;17(1):192 [FREE Full text] [doi: 10.1186/s12889-017-4065-6] [Medline: 28193194]
33. Crane TE, Parizek D, Eddy N, Franks GH, Yung AK, Merchant N, inventors, University of Arizona, assignee. Ehealth and intervention platform. United States patent application US 15/655,553. 2018 Jan 25. .
34. CentrePoint Mobile User Manual. ActiGraph. 2018. URL: https://actigraph.desk.com/customer/en/portal/articles/2515621-centrepoint-mobile-user-manual [accessed 2019-08-03]
35. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. Eur J Appl Physiol 2011 Mar;111(2):187-201 [FREE Full text] [doi: 10.1007/s00421-010-1639-8] [Medline: 20842375]
36. LaMonte MJ, Buchner DM, Rillamas-Sun E, Di C, Evenson KR, Bellettiere J, et al. Accelerometer-measured physical activity and mortality in women aged 63 to 99. J Am Geriatr Soc 2018 May;66(5):886-894 [FREE Full text] [doi: 10.1111/jgs.15201] [Medline: 29143320]
37. Sweegers MG, Boyle T, Vallance JK, Chinapaw MJ, Brug J, Aaronson NK, et al. Which cancer survivors are at risk for a physically inactive and sedentary lifestyle? Results from pooled accelerometer data of 1447 cancer survivors. Int J Behav Nutr Phys Act 2019 Aug 16;16(1):66 [FREE Full text] [doi: 10.1186/s12966-019-0820-7] [Medline: 31420000]
38. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications inc accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]
39. Duncan S, Stewart T, Mackay L, Neville J, Narayanan A, Walker C, et al. Wear-time compliance with a dual-accelerometer system for capturing 24-h behavioural profiles in children and adults. Int J Environ Res Public Health 2018 Jun 21;15(7):- [FREE Full text] [doi: 10.3390/ijerph15071296] [Medline: 29933548]

40. Tudor-Locke C, Barreira TV, Schuna JM, Mire EF, Chaput J, Fogelholm M, ISCOLE Research Group. Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the international study of childhood obesity, lifestyle and the environment (ISCOLE). Int J Behav Nutr Phys Act 2015 Mar 11;12:11 [FREE Full text] [doi: 10.1186/s12966-015-0172-x] [Medline: 25881074]
41. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
42. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc 2015 Apr;47(4):839-842. [doi: 10.1249/MSS.0000000000000476] [Medline: 25121517]
43. Loprinzi PD, Smith B. Comparison between wrist-worn and waist-worn accelerometry. J Phys Act Health 2017 Jul;14(7):539-545. [doi: 10.1123/jpah.2016-0211] [Medline: 28290761]
44. Kampshoff CS, Chinapaw MJ, Brug J, Twisk JW, Schep G, Nijziel MR, et al. Randomized controlled trial of the effects of high intensity and low-to-moderate intensity exercise on physical fitness and fatigue in cancer survivors: results of the resistance and endurance exercise after chemotherapy (REACT) study. BMC Med 2015 Oct 29;13:275 [FREE Full text] [doi: 10.1186/s12916-015-0513-2] [Medline: 26515383]
45. Miller NE, Strath SJ, Swartz AM, Cashin SE. Estimating absolute and relative physical activity intensity across age via accelerometry in adults. J Aging Phys Act 2010 Apr;18(2):158-170 [FREE Full text] [doi: 10.1123/japa.18.2.158] [Medline: 20440028]
46. Tudor-Locke C, Brashear MM, Johnson WD, Katzmarzyk PT. Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese US men and women. Int J Behav Nutr Phys Act 2010 Aug 3;7:60 [FREE Full text] [doi: 10.1186/1479-5868-7-60] [Medline: 20682057]
47. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
48. Evenson KR, Wen F, Herring AH, Di C, LaMonte MJ, Tinker LF, et al. Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91 years: the women's health initiative OPACH calibration study. Prev Med Rep 2015;2:750-756 [FREE Full text] [doi: 10.1016/j.pmedr.2015.08.021] [Medline: 26527313]
49. Trinh L, Motl RW, Roberts SA, Gibbons T, McAuley E. Estimation of physical activity intensity cut-points using accelerometry in breast cancer survivors and age-matched controls. Eur J Cancer Care (Engl) 2019 Sep;28(5):e13090. [doi: 10.1111/ecc.13090] [Medline: 31106924]
50. Behavioral Measurement and Interventions Shared Resource. University of Arizona Cancer Center. 2018. URL: https://cancercenter.arizona.edu/researchers/shared-resources/behavioral-measurement-and-interventions [accessed 2020-04-19]

Abbreviations

APAQ: Arizona Physical Activity Questionnaire
BMISR: Behavioral Measurement and Interventions Shared Resource
HIPAA: Health Insurance Portability and Accountability Act
LIvES: Lifestyle Intervention for oVarian cancer Enhanced Survival
OR: odds ratio
SOP: standard operating procedure
UACC: University of Arizona Cancer Center
USPS: United States Postal Service

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by J Brown, C Valle; comments to author 08.04.20; revised version received 03.06.20; accepted 26.07.20; published 24.09.20.

Please cite as:
Crane TE, Skiba MB, Miller A, Garcia DO, Thomson CA
Development and Evaluation of an Accelerometer-Based Protocol for Measuring Physical Activity Levels in Cancer Survivors: Development and Usability Study
JMIR Mhealth Uhealth 2020;8(9):e18491
URL: http://mhealth.jmir.org/2020/9/e18491/
doi:10.2196/18491
PMID:32969828

Original Paper

Excessive Smartphone Use and Self-Esteem Among Adults With Internet Gaming Disorder: Quantitative Survey Study

Hyunmin Kim[1,2], PhD; In Young Choi[1], PhD; Dai-Jin Kim[3], MD, PhD

[1]Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul St. Mary's Hospital, Seoul, Republic of Korea
[2]Division of Health Systems Management and Policy, School of Public Health, The University of Memphis, Memphis, TN, United States
[3]Department of Psychiatry, College of Medicine, The Catholic University of Korea, Seoul St. Mary's Hospital, Seoul, Republic of Korea

Corresponding Author:
In Young Choi, PhD
Department of Medical Informatics
College of Medicine
The Catholic University of Korea, Seoul St. Mary's Hospital
222 Banpo-daero, Seocho-gu
Seoul
Republic of Korea
Phone: 82 2258 7870
Email: iychoi@catholic.ac.kr

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2020/11/e24869/

Abstract

Background: Smartphone overuse can harm individual health and well-being. Although several studies have explored the relationship between problematic or excessive smartphone use and mental health, much less is known about effects on self-esteem, which is essential in having a healthy life, among adults with mental health disorders, including internet gaming disorder. Furthermore, given that smartphone usage differs by gender, little is known about gender differences in the relationship between smartphone overuse and self-esteem.

Objective: The objective of this study was to assess self-esteem among individuals with mental health disorders and explore the relationship with excessive smartphone use.

Methods: Participants were selected based on their responses to the internet gaming disorder assessment, which includes 9 items developed based on Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) criteria, from among a Korean cohort of smartphone users aged 20-40 years, resulting in a sample of 189 participants (men:120, women: 69). The Rosenberg self-esteem scale and the Korean smartphone addiction proneness scale were utilized to assess the outcome self-esteem with excessive smartphone use as the primary independent variable. Guided by the Bowlby attachment theory and prior studies, we selected several covariates. Generalized linear regression analyses, as well as subgroup analyses by gender, were performed.

Results: Among adults with internet gaming disorder, the average Korean smartphone addiction proneness scale score was significantly higher in women than that in men (41.30 vs. 37.94; *P*=.001), and excessive smartphone use was significantly more prevalent in women than it was in men (30.43% vs. 20.83%; *P*=.02). Our findings from the generalized linear regression analyses indicated that an increase in Korean smartphone addiction proneness scale score had a negative relationship with self-esteem among those with internet gaming disorder (β=–0.18, *P*=.001). Furthermore, our interaction models showed that, among those with internet gaming disorder, more men than women had lower self-esteem associated with an increase in Korean smartphone addiction proneness scale score and a high degree of smartphone overuse (β=–0.19, *P*=.004; β=–3.73, *P*<.001).

Conclusions: Excessive smartphone use was found to be adversely associated with self-esteem among young and middle-aged adults with internet gaming disorder; notably, more men than women were negatively influenced (regarding self-esteem) by smartphone overuse. Based on our findings, more efforts should be made to reduce excessive or problematic smartphone use by considering developing public health interventions or policy, particularly among those with mental health disorders such as internet gaming disorder.

(JMIR Mhealth Uhealth 2020;8(9):e18505) doi:10.2196/18505

KEYWORDS

excessive smartphone use; internet gaming disorder; smartphone overuse; self-esteem; mental health; gender difference; Korean smartphone addiction proneness scale; smartphone; gaming; young adult; adult; gender

Introduction

In the era of the internet and with fast-growing mobile technology availability, smartphones have become an essential part of people's daily lives because of their mobility and various functions. Without bringing a laptop everywhere, people can easily send emails, have a video conference, share files and photos, and have access to entertainment (eg, playing games, etc) by using smartphones. According to a 2015 Pew Research Center report [1], almost two-thirds of the adults in the United States own a smartphone that 46%, among them, mentioned they "couldn't live without [1]." Among other developed countries, South Korea is known for having the highest smartphone penetration and use; it was estimated that approximately 95% of South Koreans used a smartphone in 2018, which was the top rate among the 27 nations that responded to the survey [2]. However, with substantial growth in the smartphone market, a related issue—excessive or problematic smartphone use—has arisen. Evidence indicates that many people use smartphones to affect their daily lives negatively [3,4]. Constant checking of the phones and using apps all day long could exemplify excessive smartphone use, which can also result in social issues [5].

Growing evidence suggests that excessive or problematic smartphone use can adversely affect individual physical health (eg, headaches, neck, and wrist pain; obscured vision) [6,7] and mental health (eg, depression, anxiety, etc) [8-11] as well as psychological attributes such as self-esteem and self-control [10,12,13]. For instance, a systematic review by Elhai and colleagues [10] explored the relationship of psychopathology with problematic smartphone use and identified anxiety, depression, and low self-esteem as psychopathological correlates. Notably, self-esteem, defined as "the degree that a person likes, values, and approves himself or herself [14]," is essential in having a healthy life and for psychological well-being [15]. Self-esteem provides positive outcomes and benefits in life, for example, being confident in one's own abilities can help decision making and in being more resilient in coping with stress and difficulties [15]. However, evidence shows that individuals with mental health disorders such as depression, anxiety, eating disorders, and personality disorders are more likely to have low self-esteem or self-esteem deficits [16].

To explore the relationship of self-esteem with extent of excessive smartphone use, the Bowlby attachment theory [17-19] may be applied as a theoretical base. According to this theory [17-19], there are 3 attachment styles: secure, avoidant, and anxious-ambivalent. The first type, "secure attachment," can be characterized by a higher sense of self-worth, considering others trustworthy and reliable, and controlling consequences in their lives. The second type, "avoidant," is characterized as have a lower sense of self-worth and viewing others as less kind, while the third type, "anxious-ambivalent," is characterized as feeling insecure about others' responses and having a strong desire for closeness to others. Therefore, people with avoidant and anxious-ambivalent attachment styles are more likely to have negative beliefs about themselves but relatively positive views about others, which could make them more reliant on others by seeking their authorization and reassurance [20]. In this context, a mobile phone or smartphone could serve as a means for them to receive the approval and reassurance from people close to them. Evidence indicates that low self-esteem related to insecure attachment could play a role in the excessive use of mobile phones or smartphones [21,22].

Although prior research has explored relationships between excessive or problematic smartphone use and mental health or psychological attributes among general populations, much less is known about effects on self-esteem among people with mental health disorders such as internet gaming disorder [23]. Those who are addicted to online gaming tend to have specific characteristics, such as obsession, withdrawal, and loss of interest in doing other things (ie, job-related or educational activities) [24,25]. Given that low self-esteem or self-esteem deficits may be more likely to occur among individuals with mental health disorders, it is meaningful to assess self-esteem related to problematic smartphone use among these groups. Additionally, given that smartphone usage differs by gender [26,27], it is also worthwhile to examine whether there exist gender differences in the relationship between smartphone overuse and self-esteem, about which little is known. We believe that this study has contributed to filling gaps in the literature and could help in developing public health interventions and policy to reduce excessive smartphone use among adults with mental health disorders, particularly those with internet gaming disorder.

Methods

Data and Study Participants

This study included individuals recruited from various parts of South Korea (Metropolitan area, Chungcheong-do, Gyeongsang-do, Jeolla-do, and Gangwon-do/Jeju-do) from October 2018 to February 2019 through the smartphone overdependence management system. The smartphone overdependence management system is a mobile app–based system, developed in a prior study [28], to identify and examine factors associated with smartphone overdependence that can help prevent and monitor smartphone overreliance. Thus, the smartphone overdependence management system provided us with information about excessive smartphone use. In the initial research, the inclusion criteria were (1) aged 20-40 years; (2) Android -based smartphone users; and (3) smartphone usage of at least one hour or more per day. To retain participants for 4 weeks, the project utilized a specialized company to recruit and manage participants, who were randomly selected. All

participants received a small reward for their participation through the company. Since this study focused on individuals with internet gaming disorder, we utilized a questionnaire for internet gaming disorder assessment that consisted of 9 items (preoccupation, tolerance, withdrawal, persistence, escape, problems, deception, displacement, and conflict) [29] that were developed based on Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5) criteria [24]. Each of the 9 items was coded as binary (1=yes, 0=no) and summed for a total score. Using the conventional cutoff point for internet gaming disorder, we selected those with a total of 5 or above as an indication of internet gaming disorder [24] (ie, 0-4 for normal; 5-9 for internet gaming disorder), which resulted in a sample of 189 participants with internet gaming disorder (men: 120/189, 63.5%; women: 69/189, 36.5%). A survey questionnaire (written in Korean) was utilized to obtain individual-level information (social and demographic characteristics including age, marital status, education, and income; family-related characteristics such as receipt of emotional support from family and relationship with family; and adverse conditions including perceived abuse experience and experience of being bullied). The study was approved by the institutional review board of the Catholic University of Korea, Seoul St. Mary's Hospital (approval no. MC16QISI0146). All participants signed an informed consent form before the study in accordance with the Declaration of Helsinki.

Measures

Dependent Variable

The Korean version of the Rosenberg self-esteem scale [30] was used to evaluate individual *self-esteem*. This scale consisted of 10 items with a 4-point Likert scale (4=strongly agree, 3=somewhat agree, 2=somewhat disagree, and 1=strongly disagree). The example questionnaire statements for evaluating self-esteem included "I take a positive attitude toward myself," "I feel that I have many good qualities [31]." In a prior study [32], the Rosenberg self-esteem scale was found to have relatively high reliability and consistency (Cronbach α=.72). The self-esteem scale ranged between 0 and 40, in which a higher total score indicated a higher level of self-esteem.

Independent Variables

The Korean Smartphone Addiction Proneness Scale (K-SAPS), developed by the National Information Agency of Korea in 2011, was used to assess *excessive smartphone use*. The K-SAPS, which additionally includes items with unique smartphone features in the Internet Addiction Proneness Scale for Youth, was found to be highly reliable (Cronbach α=.88) in prior research [33]. The K-SAPS comprised 15 items in total with a 4-point Likert scale (4=strongly agree, 3=somewhat agree, 2=somewhat disagree, and 1=strongly disagree) and is appropriate for screening addictive smartphone use [33]. The K-SAPS consisted of 4 subdomains: disturbance of adaptive functions (5 items), virtual life orientation (2 items), withdrawal (4 items), and tolerance (4 items). Disturbance of adaptive functions was assessed, for example, with items such as "People frequently comment on my excessive smartphone use" and "Family or friends complain that I use my smartphone too much." Withdrawal, for example, was measured with items such as "It would be unpleasant if I am not allowed to use a smartphone" and "I become anxious and restless when I do not have a smartphone with me."

We included the K-SAPS as the primary independent variable in the model and also examined it as a categorical variable to assess the degree of excessive smartphone use (high, moderate, and low) based on prior research [34]. A high degree of excessive smartphone use was defined as K-SAPS≥44 or by subdomain scores ≥15 for adaptive function or ≥13 for both tolerance and withdrawal. A moderate degree of excessive smartphone use was defined as 40≤K-SAPS≤43 or any subdomain score≥14. A low degree of excessive smartphone use was defined as any not meeting these criteria.

Covariates

Guided by the Bowlby attachment theory [17-19] and empirical studies examining factors related to self-esteem [12,35,36], we selected several covariates. They included social and demographic factors (age, gender, marital status, education, and income), family-related factors (emotional support from family, relationship with the family), as well as adverse conditions (perceived abuse experience, experience of being bullied). *Age* was used as a continuous variable in this study. *Marital status*, indicating whether or not an individual was currently married, was grouped into 3 categories: never been married, married, separated/widowed. *Education*, indicating the highest level of educational attainment an individual achieved, was included as a categorical variable (less than high school, high school graduate, college graduate or above). *Income* (X Korean won or KRW; an approximate exchange rate of 2,000,000 KRW=US $1713 was applicable at the time of publication), was measured by the average monthly household income before tax categorized into X<2,000,000, 2,000,000≤X<4,000,000, 4,000,000≤X<6,000,000, and X≥6,000,000. The variable *emotional support from family* (not at all, some, a lot) was included. It was constructed based on the survey question, "How much do you think you receive emotional support from your family?" The variable *relationship with the family* (less satisfied/dissatisfied, satisfied, very/completely satisfied) was also included based on the survey question "How satisfied are you with the relationship with your family?" Additionally, we included *perceived abuse experience* and *experience of being bullied* as binary variables, assessed based on the survey questions "Have you ever felt that you received abuse from your family or someone who provided care when you were a child?" and "Have you ever experienced being bullied at school?" respectively.

Statistical Analysis

The dependent variable in this study was self-esteem, measured with the Rosenberg self-esteem scale. The independent variables were K-SAPS score and degree of excessive smartphone use (high, moderate, or low). Using descriptive statistics, we first examined the characteristics of the Korean cohort of young and middle-aged adults with internet gaming disorder and differences by gender. Specifically, we compared the extent to which men and women were different concerning individual characteristics using two-tailed independent t tests for continuous variables

and Rao-Scott chi-square tests for dichotomous variables. The threshold of .05 was used for assessing the statistical significance of those tests.

Furthermore, we utilized a generalized linear regression model to examine the association between K-SAPS and self-esteem and between the degree of excessive smartphone use and self-esteem, among adults with internet gaming disorder. The Rosenberg self-esteem scale was not normally distributed (Shapiro-Wilk normality test: $P<.001$). We conducted generalized linear regression analyses by employing 4 distinct models: Model 1 included K-SAPS score as the primary independent variable without any interaction terms. Model 2 included the degree of excessive smartphone use as the primary independent variable without any interaction terms. Model 3 included the interaction between gender and K-SAPS score. Model 4 included the interaction between gender and degree of excessive smartphone use. Notably, the interaction models were constructed based on the results from the statistical tests comparing men and women. All statistical analyses were conducted using SAS (version 9.4; SAS Institute) statistical software.

Results

Characteristics of Study Participants

Table 1 presents the characteristics of study participants by gender. The average K-SAPS (41.30 vs. 37.94; $P=.001$) and frequency of a high degree of excessive smartphone use (30.43% vs. 20.83%; $P=.02$) were significantly higher for women than those for men. Except for the K-SAPS and degree of excessive smartphone use, however, no statistically significant differences by gender were shown concerning individual characteristics. The mean Rosenberg self-esteem scale (24.17 vs. 24.07; $P=.88$) was slightly higher for men than that for women. The average age (34.45 years vs. 34.30 years; $P=.90$) for men was higher than that for women. Being married (49.28% vs. 41.67%; $P=.28$), and with higher incomes (4,000,000 earned and above) (50.72% vs. 31.67%; $P=.28$) were more common in women than they were in men. College graduate or more education (81.67% vs. 73.91%; $P=.29$) was more prevalent in men than it was in women. Receipt of emotional support from family (some and a lot) (88.41% vs. 85.83%; $P=.84$) was more common in women than it was in men, while being satisfied with the family relationship (89.17% vs. 88.41%; $P=.74$) was more prevalent in men than it was in women. Perceived abuse experiences (33.33% vs. 23.33%; $P=.13$) and incidents of being bullied at school (43.48% vs. 31.67%; $P=.10$) were more common in women than they were in men.

Table 1. Characteristics of the study participants.

Characteristic	All (N=189)	Men (n=120)	Women (n=69)	*P* value[a]
K-SAPS[b], mean (SD)	39.16 (6.91)	37.94 (7.06)	41.30 (6.13)	.001
Degree of excessive smartphone use, n (%)				.02
High (K-SAPS≥44)	46 (24.3)	25 (20.8)	21 (30.4)	
Moderate (40≤K-SAPS≤43)	52 (27.5)	28 (23.3)	24 (34.8)	
Low (K-SAPS≤39)	91 (48.2)	67 (55.8)	24 (34.8)	
Social and demographic				
Age (in years), mean (SD)	34.39 (7.9)	34.45 (8.11)	34.30 (7.61)	.90
Marital status, n (%)				.28
Never been married	102 (54.0)	69 (57.5)	33 (47.8)	
Married	84 (44.4)	50 (41.7)	34 (49.3)	
Separated/widowed	3 (1.6)	1 (0.8)	2 (2.9)	
Education, n (%)				.29
Less than high school	1 (0.5)	1 (0.8)	0 (0.0)	
High school graduate	39 (20.6)	21 (17.5)	18 (26.1)	
College graduate or above	149 (78.8)	98 (81.7)	51 (73.9)	
Income (X KRW[c] /month), n (%)				.28
X<2,000,000	33 (17.5)	23 (19.2)	10 (14.5)	
2,000,000≤X<4,000,000	77 (40.7)	53 (44.2)	24 (34.8)	
4,000,000≤X<6,000,000	45 (23.8)	24 (20.0)	21 (30.4)	
X≥6,000,000	34 (18.0)	20 (16.7)	14 (20.3)	
Emotional support from family, n (%)				.84
Not at all	25 (13.2)	17 (14.2)	8 (11.6)	
Some	123 (65.1)	78 (65.0)	45 (65.2)	
A lot	41 (21.7)	25 (20.8)	16 (23.2)	
Relationship with the family, n (%)				.74
Less satisfied/dissatisfied	21 (11.1)	13 (10.8)	8 (11.6)	
Satisfied	68 (36.0)	41 (34.2)	27 (39.1)	
Very/completely satisfied	100 (52.9)	66 (55.0)	34 (49.3)	
Perceived abuse experience, n (%)				.13
Yes	51 (27.0)	28 (23.3)	23 (33.3)	
No	138 (73.0)	92 (76.7)	46 (66.7)	
Experience of being bullied, n (%)				.10
Yes	68 (36.0)	38 (31.7)	30 (43.5)	
No	121 (64.0)	82 (68.3)	39 (56.5)	
Rosenberg Self-Esteem Scale, mean (SD)	24.13 (4.82)	24.17 (4.79)	24.07 (4.92)	.88

[a]Significance assessment of the Rao-Scott chi-square test for categorical variables and *t* test for continuous variables. The significance level of .05 was incorporated for the assessment.

[b]K-SAPS: Korean Smartphone Addiction Proneness Scale.

[c]KRW: Korean won (an approximate exchange rate of 2,000,000 KRW=US $1713 was applicable at the time of publication).

Association Between Excessive Smartphone Use and Self-Esteem Among Adults With Internet Gaming Disorder

Table 2 shows the results of our 4 generalized linear regression models for examining the estimated effects of the K-SAPS and the degree of excessive smartphone use on self-esteem among adults with internet gaming disorder. Model 1 showed an increase in the K-SAPS had a negative relationship to self-esteem (β=–0.18, *P*=.001) among adults with internet gaming disorder. Model 2 showed that compared with those with a low degree of excessive smartphone use, individuals with internet gaming disorder with a high degree of excessive smartphone use had lower self-esteem (β=–3.42, *P*<.001). Furthermore, all models showed that compared with adults with internet gaming disorder who were very/completely satisfied with their family relationship, those who were less satisfied/dissatisfied with their family relationship had lower self-esteem (model 1: β=–5.21, *P*<.001; model 2: β=–5.55, *P*<.001; model 3: β=–5.43, *P*<.001; model 4: β=–5.63, *P*<.001). Meanwhile, interactions showed that among adults with internet gaming disorder, men had lower self-esteem associated with an increase in the K-SAPS than women (β=–0.19, *P*=.004). Additionally, among adults with internet gaming disorder, men with a high degree of excessive smartphone use had lower self-esteem than women (β=–3.73, *P*<.001).

Table 2. Generalized linear models of self-esteem concerning smartphone overuse among adults with internet gaming disorder.

Variables	Self-esteem[a]							
	Model 1		Model 2		Model 3		Model 4	
	B[b]	*P* value	B	*P* value	B	*P* value	B	*P* value
K-SAPS[c]	–0.18	.001	N/A[d]	N/A	N/A	N/A	N/A	N/A
Degree of excessive smartphone use (reference: low)[e]								
Moderate	N/A	N/A	–0.13	.87	N/A	N/A	N/A	N/A
High	N/A	N/A	–3.42	<.001	N/A	N/A	N/A	N/A
Social and demographic								
Age (in years)	0.03	.52	0.03	.48	0.04	.40	0.04	.48
Gender (reference: men)	–0.66	.66	–0.51	.47	N/A	N/A	N/A	N/A
Marital status (reference: separated/widowed)								
Never been married	-6.94	.05	–5.78	.11	–6.78	.06	–5.76	.11
Married	-6.81	.05	–5.96	.08	–6.76	.05	–5.93	.09
Education (reference: college graduate or above)								
Less than high school	–4.90	.13	–4.53	.15	–4.69	.14	–4.52	.15
High school graduate	–0.94	.94	–1.18	.15	–0.58	.48	–1.14	.17
Income (X KRW[f] /month) (reference: ≥6,000,000)								
X<2,000,000	1.49	.24	1.97	.12	1.46	.24	1.99	.12
2,000,000 ≤X<4,000,000	1.09	.28	1.49	.14	1.06	.29	1.56	.13
4,000,000 ≤X<6,000,000	2.37	.02	2.87	.01	2.28	.03	2.90	.01
Family-related								
Emotional support from family (reference: a lot)								
Not at all	–0.37	.76	–0.55	.65	–0.59	.64	–0.55	.65
Some	–0.45	.61	–0.65	.45	–0.47	.60	–0.63	.47
Relationship with the family (reference: very/completely satisfied)								
Less satisfied/dissatisfied	–5.21	<.001	–5.55	<.001	–5.43	<.001	–5.63	<.001
Satisfied	–1.47	.06	–1.50	.05	–1.58	.04	–1.50	.05
Adverse conditions								
Perceived abuse experience	–0.93	.27	–1.05	.20	–1.04	.21	–1.09	.18
Experience of being bullied	–0.11	.88	–0.02	.97	–0.15	.84	0.03	.96
Interactions								
Gender×K-SAPS[g]	N/A	N/A	N/A	N/A	–0.19	.004	N/A	N/A
Gender×degree of excessive smartphone use: high	N/A	N/A	N/A	N/A	N/A	N/A	–3.73	<.001

[a]The Korean version of the Rosenberg self-esteem scale was used for the outcome assessment.

[b]B is the unstandardized coefficient.

[c]K-SAPS: Korean Smartphone Addiction Proneness Scale.

[d]N/A: not applicable.

[e]Degree of excessive smartphone use was assessed by the K-SAPS: high (K-SAPS≥44), moderate (40≤K-SAPS≤43), and low (K-SAPS≤39).

[f]KRW: Korean won (an approximate exchange rate of 2,000,000 KRW=US $1713 was applicable at the time of publication).

[g]Grand mean centering and scaling (dividing by SD) was applied to the continuous K-SAPS variable.

Discussion

Principal Findings and Implications for Public Health Interventions and Policy

To the extent of our knowledge, this study is the first, based on a Korean cohort of young and middle-aged adults with internet gaming disorder, to examine the association between excessive smartphone use and self-esteem and detect gender differences. Although several studies have investigated the relationship between problematic or excessive smartphone use and mental health [8-10,21] and psychological attributes [12,13], much less is known about the effects on self-esteem among those with mental health disorders, including internet gaming disorder. Smartphone overuse could play a role as a stimulus for gaming (by using mobile apps), which can adversely affect smartphone users' health and well-being. There may be reason for concern for those with internet gaming disorder because they may be more prone to utilizing smartphones for gaming, with features such as accessibility and mobility, which could increase online gaming dependence. In this study, we found that excessive smartphone use was adversely associated with self-esteem in adults with internet gaming disorder. Given a substantial increase in smartphone usage, developing public health interventions or programs to reduce smartphone overuse should be considered, particularly for those with mental health disorders such as internet gaming disorder.

Another finding of this study was that men were found to be more negatively influenced from excessive smartphone use concerning self-esteem than women were; notably, men with a high degree of smartphone overuse were more likely to have low self-esteem compared with women with a high degree of smartphone overuse. This finding may be explained with the circumstance that men may use their smartphones, specifically for gaming, or their gaming behavior may be carried out via their smartphone. In contrast, women may game on other internet-connected devices and use their smartphones for different functions, including social networking or information searching [27]. In other words, women may socialize more by using their smartphones or mobile phones than men do, which could more positively affect self-esteem in women than that in men. A meta-analysis by Harris and Orth [37] examined the relationship between social relationships and self-esteem with 52 longitudinal studies published between 1993 and 2016. They found a reciprocal link between self-esteem and social relationships. Specifically, positive social relationships, acceptance, and support seemed to affect self-esteem positively among individuals across all levels of development [37].

Interestingly, among adults with internet gaming disorder, those who perceived less satisfaction or dissatisfaction about their relationships with their family seemed to have lower self-esteem than those who were very/completely satisfied with their relationships with their family. Indeed, prior research identified family-related factors as predictors of self-esteem among young individuals, including family functioning and environment [38], parent-child relationship [39], and parental attitudes toward their children (perceived parental rejection and criticism and parental fairness) [40]. Furthermore, evidence suggests that parental influences on self-esteem may vary by factors such as children's age, development status, and gender. For instance, a longitudinal study [41] of 282 adolescents found that family influences (ie, perceptions about family communication) had a significant impact on self-esteem among young adolescents at Time 1; however, it appeared to not affect self-esteem among young adolescents at Time 2. This finding suggests that the effects of family factors on self-esteem may be higher among younger rather than older individuals [41]. In terms of gender in the relationship between parental influences and self-esteem, prior studies [42-44] reported varying results. For example, several reviews mentioned that family functioning and family relationships might be more influential on boys' self-esteem than on girls' self-esteem [42,43]. However, other research conducted with a sample of 230 college students found that while parental authoritativeness and authoritarianism were shown to be predictive of self-esteem, female students were more likely than male students to be affected by their parents' attitudes [44].

Limitations and Strengths

This study had some limitations. First, this study mainly examined the estimated effects of excessive smartphone use on self-esteem, and gender differences, among young and middle-aged adults with internet gaming disorder. However, evidence indicates a bidirectional relationship between addictive behaviors and psychological factors [12]. Second, due to the cross-sectional nature of the study, we may not infer a causal relationship between excessive smartphone use and self-esteem. Next, despite using measures for self-esteem and excessive smartphone use, which were found to be valid and reliable in prior research [32,33], the 4-point Likert scale (strongly agree, somewhat agree, somewhat disagree, strongly disagree) may not have been able to capture correct answers to the survey questions completely. For instance, some respondents might have intended to answer somewhere between strongly agree and somewhat agree or between strongly disagree and somewhat disagree; therefore, there may have been response bias by being limited to a scale with only 4 points [45].

Despite the limitations, there were several strengths. First, this study was conducted based on unique data, including a wide range of information about smartphone usage, internet gaming disorder, self-esteem, family-related factors, adverse experiences among young and middle-aged adults in South Korea. This allowed us to adjust for various related factors in associations between excessive smartphone use and self-esteem, which resulted in meaningful results, such as family-related factors. Second, this study, is the first (of which we are aware) to examine the relationship between excessive smartphone use and self-esteem among adults with internet gaming disorder, a population that was less explored in prior research. Finally, we further investigated gender differences in the association between smartphone overuse and self-esteem in this population, about which little is known.

Conclusions

Although smartphones, with their advantageous features, have become an essential part of people's daily lives, smartphone overuse could harm individual health and well-being. It is of

particular concern for those more susceptible or vulnerable such as individuals with mental health disorders, including internet gaming disorder. Our finding of the negative association between excessive smartphone use and self-esteem among adults with internet gaming disorder suggests that more efforts or endeavors should be made by considering developing public health interventions or programs to reduce excessive or problematic smartphone use, particularly among those with mental health disorders such as internet gaming disorder. Furthermore, our finding of gender differences in the association between smartphone overuse and self-esteem could suggest merit in developing gender-specific interventions targeting men with internet gaming disorder.

Acknowledgments

Many thanks to In Hye Yook for providing some resources and assistance for manuscript preparation and Dr. Hyun Cho and Dr. Ji-Won Chun for their involvement in data collection and management. This work was supported by the National Research Foundation of Korea (NRF) (grant nos. 2014M3C7A1062893 and NRF-2019R1A5A2027588).

Authors' Contributions

HK was involved in conceptualization, conducted data analyses, and wrote the paper. IC and DK supervised the initial project of this study and critically reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Smith A. U.S. smartphone use in 2015. Pew Research Center. 2015. URL: https://pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2020-09-19]
2. S. Korea tops smartphone penetration rate in 2018: report. Yonhap News Agency. Seoul; 2019 Feb 16. URL: https://en.yna.co.kr/view/AEN20190206001200325 [accessed 2020-03-05]
3. Cheever NA, Rosen LD, Carrier LM, Chavez A. Out of sight is not out of mind: the impact of restricting wireless mobile device use on anxiety levels among low, moderate and high users. Comput Human Behav 2014 Aug;37:290-297. [doi: 10.1016/j.chb.2014.05.002]
4. Winskel H, Kim T, Kardash L, Belic I. Smartphone use and study behavior: a Korean and Australian comparison. Heliyon 2019 Jul;5(7):e02158. [doi: 10.1016/j.heliyon.2019.e02158]
5. Rho MJ, Park J, Na E, Jeong J, Kim JK, Kim D, et al. Types of problematic smartphone use based on psychiatric symptoms. Psychiatry Res 2019 May;275:46-52. [doi: 10.1016/j.psychres.2019.02.071] [Medline: 30878856]
6. Jonsson P, Johnson PW, Hagberg M, Forsman M. Thumb joint movement and muscular activity during mobile phone texting - a methodological study. J Electromyogr Kinesiol 2011 Apr;21(2):363-370. [doi: 10.1016/j.jelekin.2010.10.007] [Medline: 21123082]
7. Lee JI, Song HS. The correlation analysis between hours of smartphone use and neck pain in the Gachon university students. The Acupuncture 2014 Jun 22;31(2):99-109. [doi: 10.13045/acupunct.2014028]
8. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. BMC Public Health 2011 Jan 31;11(1):66 [FREE Full text] [doi: 10.1186/1471-2458-11-66] [Medline: 21281471]
9. Hwang K, Yoo Y, Cho O. Smartphone overuse and upper extremity pain, anxiety, depression, and interpersonal relationships among college students. The Journal of the Korea Contents Association 2012 Oct 28;12(10):365-375. [doi: 10.5392/jkca.2012.12.10.365]
10. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017 Jan 01;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]
11. González-Bueso V, Santamaría JJ, Fernández D, Merino L, Montero E, Ribas J. Association between Internet Gaming Disorder or Pathological Video-Game Use and Comorbid Psychopathology: A Comprehensive Review. Int J Environ Res Public Health 2018 Apr 03;15(4):668 [FREE Full text] [doi: 10.3390/ijerph15040668] [Medline: 29614059]
12. Oh K, Oh G. In The Effects of Smartphone Addiction on Self-esteem and Empathy among Nursing Students. 2016 Jun 29 Presented at: International Conference on Convergence Technology, Korea Convergence Society; 2016; Jeju, South Korea.
13. Cho M. The relationships among smart phone use motivations, addiction, and self-control in nursing students. Journal of Digital Convergence 2014 May 28;12(5):311-323. [doi: 10.14400/jdc.2014.12.5.311]
14. Robinson JP, Shaver PR, Wrightsman LS, editors. Measures of Personality and Social Psychological Attitudes. San Diego: Academic Press, Inc; 1991.
15. Self-esteem check: Too low or just right? Mayo Clinic. URL: https://mayoclinic.org/healthy-lifestyle/adult-health/in-depth/self-esteem/art-20047976 [accessed 2020-03-05]

16. Silverstone PH, Salsali M. Low self-esteem and psychiatric patients: part I - the relationship between low self-esteem and psychiatric diagnosis. Ann Gen Hosp Psychiatry 2003 Feb 11;2(1):2 [FREE Full text] [doi: 10.1186/1475-2832-2-2] [Medline: 12620127]
17. Bowlby J. Attachment and Loss: Vol. 2: Separation, Anxiety and Anger. London: Hogarth Press and the Institute of Psycho-Analysis; 1973:1-429.
18. Bowlby J. Attachment and Loss: Vol. 3: Loss. London: Hogarth Press and the Institute of Psycho-Analysis; 1980.
19. Bolwby J. Attachment and Loss: Vol. 1: Attachment. London: Hogarth Press and the Institute of Psycho-Analysis; 1969.
20. Bartholomew K. Avoidance of intimacy: an attachment perspective. Journal of Social and Personal Relationships 2016 Jun 30;7(2):147-178. [doi: 10.1177/0265407590072001]
21. De-Sola Gutiérrez J, Rodríguez de Fonseca F, Rubio G. Cell-phone addiction: a review. Front Psychiatry 2016 Oct 24;7:175 [FREE Full text] [doi: 10.3389/fpsyt.2016.00175] [Medline: 27822187]
22. Billieux J. Problematic use of the mobile phone: a literature review and a pathways model. Curr Psychiatry Rev 2012 Oct 01;8(4):299-307. [doi: 10.2174/157340012803520522]
23. International Classification of Diseases 11th revision. World Health Organization. URL: https://icd.who.int/dev11/l-m/en [accessed 2020-03-05]
24. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington DC: American Psychiatric Association; 2013.
25. King DL, Delfabbro PH. The cognitive psychology of internet gaming disorder. Clin Psychol Rev 2014 Jun;34(4):298-308 [FREE Full text] [doi: 10.1016/j.cpr.2014.03.006] [Medline: 24786896]
26. Yang S, Lin C, Huang Y, Chang J. Gender differences in the association of smartphone use with the vitality and mental health of adolescent students. J Am Coll Health 2018 Oct 22;66(7):693-701. [doi: 10.1080/07448481.2018.1454930] [Medline: 29565784]
27. Pugh S. Dublin Business School. 2017. URL: http://hdl.handle.net/10788/3329 [accessed 2020-09-01]
28. Lee S, Rho M, Yook I, Park S, Jang K, Park B, et al. Design, development and implementation of a smartphone overdependence management system for the self-control of smart devices. Applied Sciences 2016 Dec 16;6(12):440. [doi: 10.3390/app6120440]
29. Lemmens JS, Valkenburg PM, Gentile DA. The Internet Gaming Disorder Scale. Psychol Assess 2015 Jun;27(2):567-582. [doi: 10.1037/pas0000062] [Medline: 25558970]
30. Jeon BJ. Self-esteem: a test of its measurability. Yonsei Nonchong 1974;11(1):107-130.
31. Rosenberg M. Society and the Adolescent Self-Image. New Jersey: Princeton University Press; 1965.
32. Rizwan M. Psychometric properties of the Rosenberg Self-esteem Scale in Pakistani late adolescents. The International Journal of Educational and Psychological Assessment 2012.
33. Kim D, Lee Y, Lee J, Nam JK, Chung Y. Development of Korean Smartphone addiction proneness scale for youth. PLoS One 2014 May 21;9(5):e97920 [FREE Full text] [doi: 10.1371/journal.pone.0097920] [Medline: 24848006]
34. Jo H, Na E, Kim D. The relationship between smartphone addiction predisposition and impulsivity among Korean smartphone users. 2017 Apr 23;26(1):77-84. [doi: 10.1080/16066359.2017.1312356]
35. Han SS, Kim KM. [Influencing factors on self-esteem in adolescents]. [Taehan Kanho Hakhoe Chi] J Korean Acad Nurs 2006 Feb;36(1):37-44. [doi: 10.4040/jkan.2006.36.1.37] [Medline: 16520562]
36. McClure AC, Tanski SE, Kingsbury J, Gerrard M, Sargent JD. Characteristics associated with low self-esteem among US adolescents. Acad Pediatr 2010 Jul;10(4):238-244 [FREE Full text] [doi: 10.1016/j.acap.2010.03.007] [Medline: 20605547]
37. Harris MA, Orth U. The link between self-esteem and social relationships: a meta-analysis of longitudinal studies. J Pers Soc Psychol 2019 Sep 26. [doi: 10.1037/pspp0000265] [Medline: 31556680]
38. Lian, Tam C, Fatimah Y. The effects of family functioning on self-esteem of children. European Journal of Social Sciences 2009;9(4):643-650.
39. Amato PR. Marital conflict, the parent-child relationship and child self-esteem. Family Relations 1986 Jul;35(3):403-410. [doi: 10.2307/584368]
40. Openshaw DK, Thomas DL, Rollins BC. Parental influences of adolescent self-esteem. The Journal of Early Adolescence 2016 Jul 26;4(3):259-274. [doi: 10.1177/0272431684043010]
41. Heaven PC. Perceptions of family influences, self-esteem and psychoticism: a two-year longitudinal analysis. Personality and Individual Differences 1997 Oct;23(4):569-574. [doi: 10.1016/s0191-8869(97)00084-6]
42. Demo DH, Small SA, Savin-Williams RC. Family relations and the self-esteem of adolescents and their parents. Journal of Marriage and the Family 1987 Nov;49(4):705-715. [doi: 10.2307/351965]
43. Holmbeck G, Hill J. A path-analytic approach to the relations between parental traits and acceptance and adolescent adjustment. Sex Roles 1986 Mar;14:315-334. [doi: 10.1007/bf00287582]
44. Bun JR, Louiselle PA, Misukanis TM, Mueller RA. Effects of parental authoritarianism and authoritativeness on self-esteem. Pers Soc Psychol Bull 1988 Jun 02;14(2):271-282. [doi: 10.1177/0146167288142006] [Medline: 30045478]
45. Baron H. Strengths and limitations of ipsative measurement. Journal of Occupational and Organizational Psychology 1996;69(1):49-56. [doi: 10.1111/j.2044-8325.1996.tb00599.x]

Abbreviations

KRW: Korean won
K-SAPS: Korean smartphone addiction proneness scale

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by E Ennis, S Byrne, J Chen-Sankey; comments to author 29.06.20; revised version received 10.08.20; accepted 13.08.20; published 29.09.20.

Please cite as:
Kim H, Choi IY, Kim DJ
Excessive Smartphone Use and Self-Esteem Among Adults With Internet Gaming Disorder: Quantitative Survey Study
JMIR Mhealth Uhealth 2020;8(9):e18505
URL: https://mhealth.jmir.org/2020/9/e18505
doi:10.2196/18505
PMID:32990637

Original Paper

Supervised Digital Neuropsychological Tests for Cognitive Decline in Older Adults: Usability and Clinical Validity Study

Francesca Lunardini[1*], PhD; Matteo Luperto[2*], PhD; Marta Romeo[3], MSc; Nicola Basilico[2], PhD; Katia Daniele[4], MSc; Domenico Azzolino[4,5], MSc; Sarah Damanti[4,5], MD; Carlo Abbate[4], PhD; Daniela Mari[4], MD; Matteo Cesari[4,6], MD; Nunzio Alberto Borghese[2], PhD; Simona Ferrante[1], PhD

[1]Nearlab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy

[2]AISLab, Department of Computer Science, University of Milano, Milano, Italy

[3]School of Computer Science, The University of Manchester, Manchester, United Kingdom

[4]Geriatric Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy

[5]Department of Nutritional Sciences, University of Milano, Milano, Italy

[6]Department of Clinical and Community Sciences, University of Milano, Milano, Italy

[*]these authors contributed equally

Corresponding Author:
Francesca Lunardini, PhD
Nearlab
Department of Electronics, Information and Bioengineering
Politecnico di Milano
Via Giuseppe Colombo 40
Milano, 20133
Italy
Phone: 39 3334335004
Email: francesca.lunardini@polimi.it

Abstract

Background: Dementia is a major and growing health problem, and early diagnosis is key to its management.

Objective: With the ultimate goal of providing a monitoring tool that could be used to support the screening for cognitive decline, this study aims to develop a supervised, digitized version of 2 neuropsychological tests: Trail Making Test and Bells Test. The system consists of a web app that implements a tablet-based version of the tests and consists of an innovative vocal assistant that acts as the virtual supervisor for the execution of the test. A replay functionality is added to allow inspection of the user's performance after test completion.

Methods: To deploy the system in a nonsupervised environment, extensive functional testing of the platform was conducted, together with a validation of the tablet-based tests. Such validation had the two-fold aim of evaluating system usability and acceptance and investigating the concurrent validity of computerized assessment compared with the corresponding paper-and-pencil counterparts.

Results: The results obtained from 83 older adults showed high system acceptance, despite the patients' low familiarity with technology. The system software was successfully validated. A concurrent validation of the system reported good ability of the digitized tests to retain the same predictive power of the corresponding paper-based tests.

Conclusions: Altogether, the positive results pave the way for the deployment of the system to a nonsupervised environment, thus representing a potential efficacious and ecological solution to support clinicians in the identification of early signs of cognitive decline.

(JMIR Mhealth Uhealth 2020;8(9):e17963) doi:10.2196/17963

KEYWORDS

aging; Bells Test; computerized testing; dementia; early diagnosis; eHealth; mild cognitive impairment; neuropsychological assessment; Trail Making Test

Introduction

Background

In the near future, the exponential growth of the number of people 65 years or older is expected to have an impact on health care systems because of the physical and cognitive decline typically associated with the aging process. In terms of cognitive decline, population aging goes hand in hand with the rapid increase in the number of people with dementia and the related increase in public costs. Indeed, as reported by the World Health Organization, the total estimated worldwide cost of dementia in 2010 reached US $604 billion, of which approximately one-third was spent in Western Europe [1].

The transient step between physiological aging and dementia is known as mild cognitive impairment (MCI); MCI is a condition characterized by cognitive weakening that is not yet producing a clinically significant effect on daily activities but can be detected through clinical examinations or formal cognitive tests [2,3]. Currently, there is no effective treatment for dementia; however, nonclinical interventions such as cognitive training started in the MCI stage [4] can delay the onset of dementia and extend the duration of independent living [5]. Therefore, early diagnosis is crucial.

Currently, neuropsychological assessment represents an important tool for the diagnosis of dementia and MCI. It consists of a set of multi-item rating scales and batteries of brief cognitive tests that evaluate the different cognitive functions and is administered by a specialist in controlled environments, typically hospitals or clinical facilities [6]. The controlled environment setting might easily delay the diagnosis [7] for 2 main reasons: (1) the long waiting times of outpatient facilities [8] and (2) the fact that individuals are often examined after the manifestation of symptoms and when serious concerns are raised by their referents. In addition to this delayed diagnosis, a recent meta-analysis [9] showed that the proportion of undetected dementia is above 60%.

In this framework, the American Psychological Association recognized the importance of computerized testing [10] as an essential part of the screening procedures in the nearest future. Indeed, computerized assessment has potential advantages over the widely used paper-and-pencil testing, including cost and time efficiency, accurate recording of responses, automatic extraction of quantitative features related to test performance, computation of additional fine indicators, and comparison of the patient's outcome between different sessions over time. Owing to the pervasiveness of contemporary computing technologies, computerized testing can be a solution to counter delayed diagnosis of dementia, eventually allowing the adoption of proper preventive measures; such tools can be used in clinical facilities to increase the effectiveness of medical services or be deployed in home environments to detect possible cognitive impairment earlier than sporadic medical visits.

Prior Work

Although initial attempts to introduce computerized versions of the classic paper-and-pencil neuropsychological tests have been reported [11-16], the achievement of feasible, effective, and ecological computerized testing is hampered by 3 main factors: (1) the lack of normative data to support tool validity, (2) technology use anxiety of older adults, and (3) the challenging implementation of test supervision. Concerning the first challenge, Zygouris and Tsolaki [17] pointed out that in most cases data supporting the concurrent validity of computerized testing with formal paper-and-pencil tests are largely incomplete or inconclusive. Against this backdrop, 2 computerized test batteries—the CNS Vital Signs (CNSVS; CNS Vital Signs, LLC) [18] and the Cambridge Neuropsychological Test Automated Battery (CAN-TAB; Cambridge Cognition Ltd) [19]—were particularly successful in providing normative databases. The CNSVS [18] is self-administered and implements a number of heterogeneous tests, including finger tapping, verbal and visual memory, test of shifting attention, digit symbol coding, Stroop test, and continuous performance test. The CAN-TAB [19] is administered by a trained technician and includes various tests that assess visual memory, executive function, attention, semantic and verbal memory, decision making, response control, and social cognition that can be combined into different batteries.

For the second challenge, the reduced familiarity with technology that characterizes older adults can affect both their performance and willingness to undergo computerized testing. Werner and Korczyn [20] conducted interviews to examine factors associated with the expressed willingness to use computerized systems to diagnose dementia and reported a strong effect of technology anxiety, particularly for participants with lower socioeconomic status and for female users. In this framework, NeuroTrax Mindstreams (NeuroTrax Corporation) [16] was reported to be particularly user-friendly [21]. Mindstreams implements digitized adaptations of tests designed to study different domains. Tests are performed with the support of an examiner who is needed throughout the testing process and include assessing verbal and nonverbal memory, visual-spatial skills, verbal fluency, information processing, attention, executive function, and motor skills. In a study by Fillit et al [22], the battery was rated as easy to use by most older adults, even by those with significant cognitive impairment.

Concerning the third challenge, despite the great achievements of state-of-the-art computerized testing systems, automatizing test supervision remains to be a crucial problem. Usually, computerized testing is either performed under the supervision of an expert professional (as requested for the CAN-TAB and Mindstreams batteries) or self-administered (as for the CNSVS). In the first case, home-deployment of computerized testing is hindered by the need for a professional, whereas in the latter case, the absence of supervision to guarantee and guide proper test execution in uncontrolled environments may affect test results [17].

Objectives

In this study, we developed a platform that provides a computerized version of 2 neuropsychological tests commonly used to assess dementia and MCI, the Trail Making Test (TMT) [23] and the Bells Test [24]. The goal was to go one step further

to present state-of-the-art systems and provide a tool that could effectively support the diagnosis of cognitive decline. Our system provides test supervision through an intelligent vocal assistant (VA) embedded in the platform and introduces a replay functionality to allow remote inspection of the user's performance after test completion. Such innovative functionalities potentially allow the delivery of the system directly to the user's home or for usage in clinical facilities without strict supervision to support current medical services. To adopt the system in nonsupervised environments, extensive functional testing of the platform was performed, together with a clinical validation of the computerized version of the tests. The validation had the two-fold aim of evaluating the system's usability and acceptance and investigating the concurrent validity of the computerized assessment compared with the corresponding paper-and-pencil tests in identifying clinical cognitive decline. This paper presents the results obtained from a study on 83 older adults.

Methods

Design Approach and Definition of Requirements

Batteries of neuropsychological tests are usually administered by a neuropsychologist at clinical premises. The neuropsychologist explains the test to the user, supervises test execution (possibly notifying and correcting the user in case of errors), and assigns the score after test completion. The level of supervision is strictly dependent on each test protocol.

To design the supervised digitized neuropsychological tests, the principles of the design thinking process were adopted [25]. First, brainstorming sessions with clinical experts (neuropsychologists and geriatricians) and technical partners were organized to identify user and functional needs. The following requirements were obtained:

- Test supervision should mimic the one offered by the neuropsychologist during the paper-based version of the test. To this aim, events requiring supervision should be promptly and consistently recognized to trigger the related intervention.
- Test raw data and indicators should be stored and accessible for evaluation by authorized experts. Pseudonymization of data should be provided.
- A replay functionality—capable of recording a test session, reproducing it, and highlighting relevant events—should be provided to allow clinicians to remotely inspect the test after its execution.
- The digitized version of the test should be conveyed by standard consumer technology.
- The digitized version of the test should be intuitive and allow easy interactions with the user (as tests are designed to be performed by patients with cognitive decline).
- Test scores should not depend on any speech-based interaction with the user, which should aid in the execution of the actual test.

The prototyping process consisted of different testing iterations to refine the requirements and, consequently, system development. Clinicians and target users were involved during these early stages of design. The first testing iterations consisted of showing the software to a team of geriatricians and neuropsychologists. Once the software was collectively approved, some tests on older adults were performed to check the software behavior with the target users and to gather additional feedback for usability improvement. The feedback from 4 subjects recruited from a senior association was leveraged to improve the system and fine-tune the interface and the interfacing modalities (eg, implementation of palm rejection, inclusion of additional written indications to facilitate test comprehension and execution, implementation of specific software improvements to minimize errors in the automatic scoring). After these refinements, testing was conducted on a large population of users.

Test Selection

On the basis of the identified functional requirement and the experience of clinicians, the TMT including its 2 conditions, Part A (TMT-A) and Part B (TMT-B), and the Bells Test were chosen.

The TMT was chosen as it is a well-established predictor of cognitive functional abilities in both healthy older adults and patients with MCI [23]. In the TMT-A, subjects are asked to draw a line connecting, in sequential order, 25 target numbers presented on an A4 paper sheet (portrait orientation). The TMT-B requires subjects to connect, in sequential and alternate order, 13 target numbers and 12 target letters (ie, 1, A, 2, B...N, and 13). Both tests must be executed as quickly as possible, possibly without lifting the pen from the paper. If patients make an error, the examiner reorients them to the last correct target. The test ends when the correct sequence is completed with the last target, and the main outcome is the time of completion. The 2 versions of the test allow the examiner to assess different cognitive domains: psychomotor speed and tracking for the TMT-A and processing speed, mental flexibility, and executive functions for the TMT-B. Furthermore, to better isolate the cognitive processes associated with the TMT-B performance, it is common practice to subtract the time required to complete the TMT-A from the time required to complete the TMT-B, thus deriving a new parameter (TMT-BA) [26].

The Bells Test was selected to investigate attentional functions through a visual search task. In the version proposed by Gauthier et al [24], subjects are asked to find and mark 35 targets (black-ink drawings of bells), presented on an A4 paper sheet (landscape orientation), among 280 distractors. If the subject stops before all the bells are encircled, the examiner gives one incitement to verify that all the targets have been found. After this incitement, the test ends when subjects stop their searching activity. The main outcome of the test is the number of targets correctly identified.

Supervised Computerized Neuropsychological Tests

System Architecture

The supervised computerized neuropsychological tests platform presents a hierarchical architecture (Figure 1) based on the integration of 2 main components: an app running on a tablet that implements the digitized tests (TabletWebApp) and an intelligent VA that provides the supervision required to conduct

the test autonomously in an uncontrolled environment. The web app, TabletWebApp, displays the tests on a tablet, processes in real time the trace executed by the user, recognizes relevant events such as users' errors, and interacts with the VA during test execution. The VA, owing to its interactions with the TabletWebApp during test execution, interactively explains the test to the user, supervises test progression, and provides adequate feedback to the user through speech interventions. The integration between the 2 main components is enabled by a bidirectional communication channel following the Message Queue Telemetry Transport (MQTT) protocol, a publish or subscribe transport layer built on top of TCP/IP (Transmission Control Protocol and Internet Protocol) widely used for internet of things applications. The messages sent and received via this channel allow the TabletWebApp and VA to exchange information about the present state of the test, so that consistent speech interventions can be triggered.

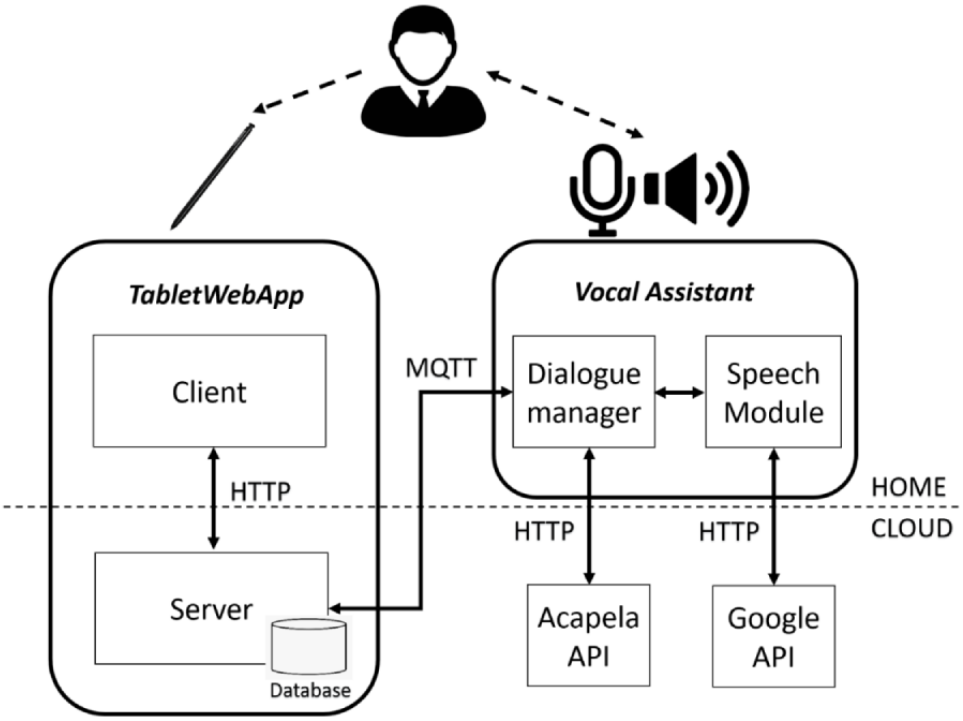
The TabletWebApp comprises a server component that resides in the cloud and a client component embodied in the tablet (Figure 1). A Samsung Galaxy Tab A6 10.1 with S Pen (115 mm long stylus, with a squared section of 6.5 mm×5 mm) was used, thus making it the most similar solution to the paper-and-pencil approach. The client displays, at a frame rate of 30 Hz, the current view of the test (as received from the server) along with the trajectory of the stylus to provide the user with the sensation of drawing on the screen. It also logs the current stylus position and communicates it to the server in real time. The client features a user front end based on HTML5 and JavaScript with graphical user interfaces (GUIs). The client interfaces have been developed to work on any touch device to be adaptable to the users' needs; the test canvas is displayed on a full screen to avoid distraction. The server component is based on Node.js; it keeps track of the present state of the test and provides the app's view via http to the tablet client. To secure communication, the encrypted http secure protocol is leveraged between the client and server. All raw data (along with the indicators and parameters are presented in the Digitized TMT and Bells Test section) are stored server-side inside a NoSQL (Not only Structured Query Language) database.

As for the VA, a voice user interface was chosen to meet the requirement of an easy and intuitive interaction with the user. Previous work [27] on senior users showed that a voice-based interaction increases the overall engagement with the device, compared with the alternative of operating a GUI. The VA interacts with users through speech, adopting a 2-phase implementation: (1) speech recognition and (2) speech synthesis. The VA includes a speech module and a dialogue manager. The speech module encompasses a speech recognition service that exploits the speech-to-text Google Cloud Platform Speech Application Programming Interface (API). The dialogue manager implements a high-level logic for speech interaction, supervises speech by running a Finite-state Machine that tracks the present state of the dialogue engaged with the user, manages the speech synthesis based on Acapela Voice as a Service, and manages the interaction with TabletWebApp via MQTT.

The dialogue manager calls the speech module whenever an utterance from the user is needed for the correct flow of the interaction. The speech module is responsible for opening the microphone to listen to the user's utterances, sending the recorded utterance to the Google Cloud Platform for recognition, and analyzing the text returned from the Google Speech API. To provide the most natural interaction, the VA has been designed to recognize different utterances that provide the same semantic meaning. For this reason, the inputs detected by the microphone and processed by the Google API are further analyzed to look for relevant keywords. The speech from the user is sent to the Google Speech API which returns a text with the transcription of the user's speech. In this text, the speech module looks for relevant keywords that are organized in sets of different input dictionaries so that different utterances with the same semantic meaning are clustered in a single semantic input (eg, "Yes, I understood," "I have understood," and "I understand" are associated to the same semantic input "understood"). Once a relevant keyword is found, it is considered as the user's answer and is communicated to the dialogue manager. The dialogue manager then selects the next correct utterance of the VA from a set of output dictionaries and reproduces the associated audio file obtained through the Acapela API. Each output dictionary is appropriate for the specific state of the dialogue manager, which is changed after receiving the answer from the speech module. Once a transition between the 2 states of the dialogue manager is executed, a message is sent to the TabletWebApp, which properly reacts to work in parallel with the VA. The state transitions of the dialogue manager can also be activated by the TabletWebApp, which automatically recognizes the user's actions and sends a message to trigger proper transition.

As described by its architecture, the platform is distributed across a local client, and a server is deployed in the cloud. Consequently, all the described functionalities constantly require a reliable internet connection.

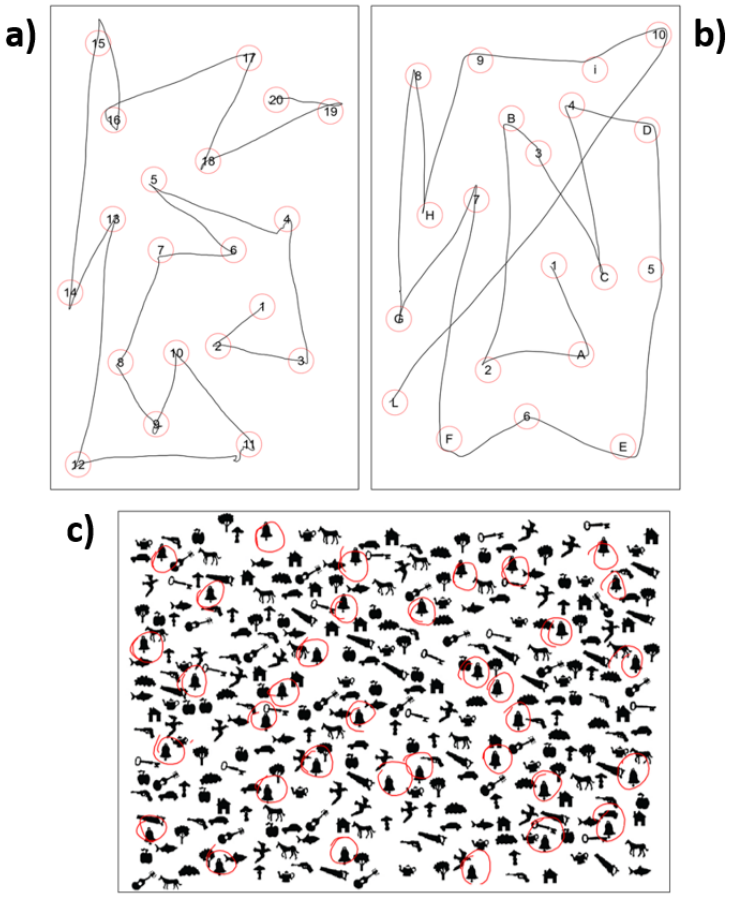
Figure 1. System architecture. API: application programming interface; HTTP: Hypertext Transfer Protocol; MQTT: Message Queue Telemetry Transport.



Digitized TMT and Bells Test

The digitized TMTs were designed to be structurally similar to the paper-based version proposed by Giovagnoli et al [28]; however, given the reduced dimensions of the tablet (14.8 cm × 21 cm) compared with the A4 paper sheet, the number of targets was decreased to 20 for both versions (TMT-A: 1-20; TMT-B [Italian version]: 1A-10L; and TMT-B [English version]: 1A-10J; Figure 2) [13]. For both TMT-A and TMT-B, the first and last targets were indicated by a written sign. In the digitized TMT, a target is considered reached when the stylus trace hits the target active area, which corresponds to the circle shown on the screen. To increase the robustness of the platform, the active area of the last target is increased by 20% when it is next to be reached according to the correct order. This choice was made because, as noticed from pretesting on potential users (see the *Design Approach and Definition of Requirements* section), it may occur that at the end of the test, the pen trace approaches the final target without really entering the circle. The software logs the events (target hit), analyzes the connection order, and classifies each reached target as a correct target (respecting the order), a repeated target (previously hit in the sequence), or an error (violating the order). The TMT starts automatically when the first target is entered and ends automatically when the last target is reached after being connected in the correct order with all previous ones or after a 5-min timeout.

Figure 2. Layout of the digitized tests: (a) Trail Making Test (TMT)-Part A, (b) TMT-B, and (c) Bells Test.



The layout of the digitized Bells Test, on the other hand, was identical to that proposed by Vallar et al (35 targets and 280 distractors) [29]; however, all graphical elements were scaled to fit the tablet size (Figure 2). Rather than keeping the symbol size constant and reducing the number of symbols compared with the paper-based version, we opted for this solution to maintain the original configuration of the Bells Test with the characteristic distribution and ratio between targets and distractors. Moreover, pretesting on potential users (see the *Design Approach and Definition of Requirements* section) reported that the symbol size on the tablet was not considered problematic for test execution. In the digitized Bells Test, a target is considered as found when a collision occurs between the stylus trace and the target active area. The target's active area here is a circle centered in the symbol centroid, with an area that exceeds the symbol surface by 20%. The software logs the events (target found), analyzes the symbol marked, and classifies it as a correct target (bell) or as an error. The Bells Test ends when all the bells have been correctly identified, when the user believes to have found all the bells and declares it, or after a 5-min timeout.

For both tests, the test layout data (target coordinates) and the collected raw data (stylus coordinates at each time instant) are stored in a cloud database, together with a log of events that occurred during test execution. Each detected event is described in the log by the following fields: target ID, timestamp, 2-dimensional pen coordinates (if present), event type (target-TMT: if the correct target was connected; target-Bells: if a bell was correctly identified; error-TMT: if a target was connected in the wrong order; error-Bells: if a symbol different than a bell was identified; and repetition-TMT: if a target already correctly connected in the sequence was repeated).

The stored data, anonymized in compliance with users' privacy, are used with a two-fold purpose: (1) computing test indicators and (2) providing a remote replay functionality.

To compute test indicators for the 2 tests, the following evaluation indicators are chosen:

- TMT:
  - Time: time required to complete the test (main outcome of the test).
  - Targets: total number of target-TMT events.
  - Errors: number of error-TMT events.
  - Errors/targets the number of error-TMT events divided by the number of target-TMT events.
  - ΔT: the average time between 2 successive target-TMT events.
  - ΔT_n: the time between 2 successive target-TMT events, averaged over the last 21−n target-TMT events (with 1≤n≤20).
- Bells:
  - Targets: total number of target-Bells events (main outcome of the test).
  - ΔT: the average time between 2 successive target-Bells events.

- ΔT_n: the time between 2 successive target-Bells events, averaged over the last 36−n target-Bells events (with 1≤n≤35).
- NIDmin: The number of inversions in direction (NID) was computed as a measure of the regularity of the patient's scanning strategy. The *scanning* trace of the Bells Test execution was calculated by connecting the coordinates of the first 25 target bells in the order of detection and then decomposed into horizontal (x) and vertical (y) components. The x- and y-scanning traces were smoothed (moving average over a span of 3 elements). For each of the 2 directions, the NID was computed as the number of zero-crossings of the derivative of the smoothed trace. As the scanning strategy can be organized either in a vertical or horizontal pattern, the smallest NID over the 2 directions was retained.
- 1Bell: the location of the first target-Bells event, expressed in terms of subarea. The canvas of the test is divided into 9 rectangular subareas of equal sizes (3 rows [1, 2, and 3]×3 columns [A, B, and C]).

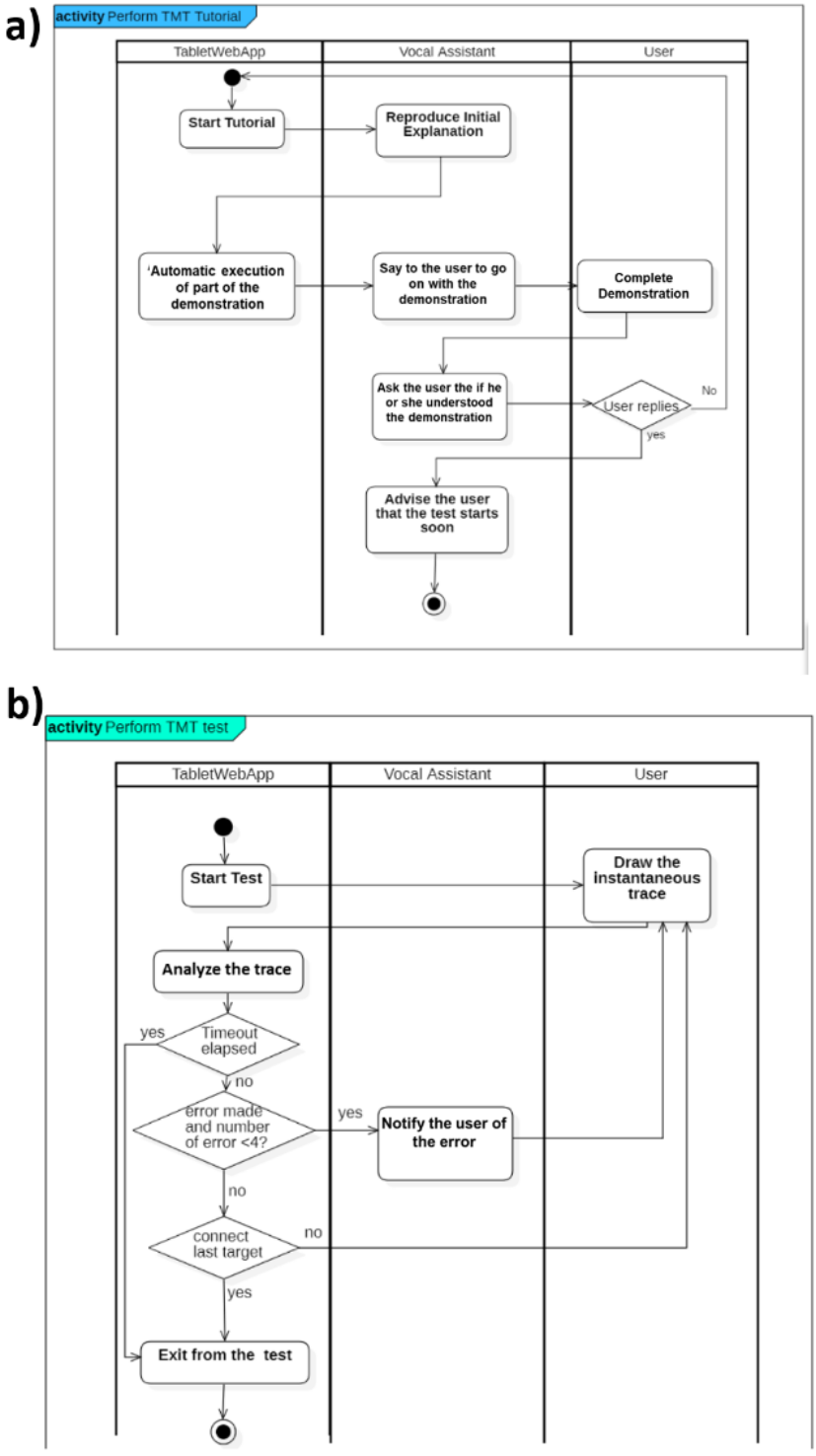
Regarding the second aim, the test data are used to provide a remote replay functionality, which reproduces the test and highlights all the events that had been detected by the software. The aim of the replay is to provide clinicians and caregivers with the possibility of watching a deferred session of the test through any web browser.

Test Progression

The session begins when the user starts the test app by tapping its icon on the tablet's home screen. First, a short explanation is presented by the VA. Once the user acknowledges the comprehension of the explanation, the tutorial of the digitized test starts. Such a tutorial aims to replicate the one conducted by the neuropsychologist during the paper-based test and consists of a short example of the test: a small subset of targets is presented to the user, who hears the VA explaining the test while the app performs the initial part of the demonstration by highlighting the first targets; at the end, the VA asks the user to practice with the remaining ones. The tutorial flowchart for the TMT is shown in Figure 3. Once the tutorial is over, the actual test starts. For the TMT, as soon as the user fails to connect targets in the right order, an error is detected by the tablet app, and an error message is shown on the tablet screen. At this point, virtual supervision is triggered: the VA notifies the user that an error has been made and that test execution should be resumed from the last target correctly identified. As in the paper-based test, the user receives notifications for a maximum of 3 errors. The high-level activity diagram of the TMT execution is shown in Figure 3. During the Bells Test, supervision is implemented under the following 2 conditions: (1) if no additional bells are identified for more than 45 seconds, the VA asks the users if they believe to have found all the targets, and the test is either resumed or terminated based on the users' answer, or (2) if the users declare to have finished the test, the VA recognizes the command and the test is terminated.

TMT-A and TMT-B are conducted in a row, and, at the end of both tests, users are asked for their availability to take the Bells Test. At the end of the whole assessment, the VA reproduces a general accomplishment notification to the user (eg, "Thank you for taking part in this test"). Such a general remark is provided regardless of the actual results of the tests, as requested by clinical partners following the paper-and-pencil evaluation.

Figure 3. Unified Modeling Language activity diagram for the (a) Trail Making Test tutorial and (b) test execution.



System Testing and Validation

In this section, we report a series of tests conducted on the devised system. First, functional testing was executed to verify the robustness of the entire platform (see the *System Architecture Functional Testing* section). In addition, validity and usability of the tablet-based neuropsychological testing were evaluated for potential users (see the *Clinical Validation of the Digitized Tests section*).

System Architecture Functional Testing

The 2 components of the proposed framework, namely the TabletWebApp and the VA, were extensively evaluated in terms of robustness and functionality both as stand-alone components and as an integrated system.

Functional Evaluation of VAs

Supervision over test execution is performed by a series of speech-based interactions between the VA and the user. Test

supervision mimics the one performed by clinicians at clinical premises and consists of a series of sentences with a very simple structure. Speech recognition is of paramount importance as issues related to the understanding of the users' answers may undermine the effectiveness and usability of the entire platform, resulting in misuse of the system. The correct identification of the semantic meaning of the user's utterances and the correct interaction flow were assessed in a controlled setting (ie, university laboratory) on adult subjects. No constraint on the age of the users was imposed as the subject's age should not affect the speech recognition performance of the Google Cloud Platform Speech APIs. Subjects were instructed to interact with the system without any specific constraint to test the robustness of recognizing natural language variants. Possible noise sources that could undermine the performance of VA speech recognition were simulated (eg, background noise coming from a television and people talking in the background). Overall, 200 tests were performed.

Functional Evaluation of the Integrated System

To test the communication and integration of the 2 main components (TabletWebApp and VA), a set of 15 experimental runs with all possible branches of the interaction between the user, VA, and the TabletWebApp were performed. From the data collected, a detailed analysis of the workflow steps was performed by computing the following metrics: (1) passed steps, that is, steps successfully completed with their expected outcome; (2) incomplete steps, that is, steps not completed because of adverse events/errors, whose possible occurrence was expected and consequently managed by the system (eg, MQTT disconnection); (3) blocked steps, that is, steps not started because of incomplete previous steps; and (4) failed steps, that is, steps resulting in unexpected outcomes, which were not managed by the system.

Clinical Validation of the Digitized Tests

Testing the validity and usability of the tablet-based TMT and Bells Test is necessary before a possible deployment of the full system guided by a VA in an unsupervised environment. For this reason, the digitized tests were performed on a group of older adults in clinical premises. During the clinical validation of the digitized tests, the experimental sessions were conducted under the supervision of a clinician specifically trained to use the tablet app.

Participants

With the aim of improving the early assessment of cognitive decline, this study addresses older adults who do not have high cognitive impairment. For this reason, the following inclusion criteria were defined: (1) aged ≥65 years and (2) a Mini-Mental State Examination score ≥20. Participants were recruited by neuropsychologists of the geriatric unit of the Foundation IRCCS (Istituti di Ricovero e Cura a Carattere Scientifico) Ca' Granda Ospedale Maggiore Policlinico (Milan, Italy) from among the patients undergoing a neuropsychological visit that included the administration of a battery of standard neuropsychological tests, including the original paper-based TMT and Bells Test.

The ethical board of the Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico approved the study protocol (October 12, 2017; n°642_2017bis).

Experimental Protocol

To reduce the effect of possible facilitation of the computerized test execution, each participant started the battery of standard neuropsychological tests with the paper-and-pencil TMT and Bells Test. On the basis of the outcome of the visit, the neuropsychologist provided a diagnosis and the subject was categorized into 1 of the following 3 groups: *normal*, *MCI*, or *dementia*.

The administration of the digitized tests took place after the completion of the clinical neuropsychological assessment so that the computerized tests were performed about 2 hours after the paper-based tests. Participants were seated at a table with the tablet flat in front of them. In case of poor eyesight, participants were instructed to wear their glasses. The digitized neuropsychological tests were performed under the supervision of a trained clinician in the following order: (1) TMT-A, (2) TMT-B, and (3) Bells Test. Before each digitized test, subjects were provided with an oral explanation of the test and were allowed to practice the tutorial on the tablet. Each digitized test was started only after the participant had shown a proper understanding of the execution mechanism.

After the completion of the computerized neuropsychological tests, subjects were asked to fill out a questionnaire. The questionnaire consisted of 2 parts:

- A series of 7 questions were related to the ease of use of the digital platform, the clarity of the provided explanations and the graphical interface, the overall satisfaction, and the user's familiarity with technology. The following questions were rated on a 3-point Likert scale:
  - I believe the test explanation and tutorial were clear.
  - I have found the graphical interface of the TMT to be simple and intuitive.
  - I have found the graphical interface of the Bells Test to be simple and intuitive.
  - I am satisfied with the overall experience.

  The following were yes or no questions:
  - Have you ever used a computer?
  - Have you ever used a tablet?
  - Have you ever used a smartphone?

- The Italian version of the System Usability Scale (SUS) questionnaire [30].

Once the experimental session was completed, a neuropsychologist (who was not present during the execution of the digitized tests) remotely watched users' performances and scored the computerized tests observed through the replay functionality.

Data Analysis and Statistics

Data analysis was conducted with 4 main aims to (1) test system acceptance and usability, (2) test system validity, (3) study and compare the discriminative power of the paper-based and

digitized test outcomes, and (4) study the digital-specific indicators in the different diagnostic groups.

System Acceptance and Usability

For each diagnosis group separately (dementia, MCI, normal) a frequency analysis (to understand how often each answer was chosen within the same group) was conducted on the answers to the items of the questionnaires was carried out. As for the SUS, the original scores were converted into a 0 to 100 range by following the guidelines provided in a study by Brooke [30]. An SUS score above 68 was considered above average.

System Validity

System validity was assessed by comparing the scores assigned to the digitized test by the software with both the performance assessed by a neuropsychologist through the replay functionality (see the *Digitized Test Versus Replay* section) and the score obtained in the paper-based counterpart (see the *Digitized Test Versus Paper-Based Test* section).

- Digitized Test Versus Replay
  In the TMT, the sporadic occurrence of the error-TMT events resulted in zero-clustered data that were not well suited for a correlation analysis [31]; therefore, the sensitivity, specificity, and accuracy of the computerized tests were computed according to the following definitions: true-positive is the number of error-TMT events correctly identified; true-negative is the number of target-TMT events correctly identified; false-positive is the number of events erroneously identified as error-TMT; and false-negative is the number of error-TMT events not identified.
  For the Bells Test, the Spearman correlation between the outcome of the digitized tests automatically computed by the TabletWebApp (targets) and the score assigned by a clinician to the execution of the same test inspected through the replay functionality was analyzed.
- Digitized Test Versus Paper-Based Test
  Concurrent validity of the digitized neuropsychological tests with regard to the paper-based tests was tested by running a Spearman correlation between the primary outcomes (time for TMT-A, TMT-B, and TMT-BA and targets for Bells Test) of the digitized and paper-based tests. In cases of a weak correlation, we investigated whether 1 of the 2 versions of the test systematically returned a higher number of targets through a Kruskal-Wallis test.

Discriminative Power of Neuropsychological Test Outcomes

For both paper-based and digitized versions, we studied the possible differences in test outcomes (time for TMT-A, TMT-B, and TMT-BA and targets for Bells Test) based on the subjects' neuropsychological diagnoses (*normal*, *MCI*, and *dementia*). The Kruskal-Wallis test with a Bonferroni adjustment for pairwise comparison was performed.

Discriminative Power of Novel Indicators Derived From the Digitized Tests

To study the ability of the digital-specific indicators to distinguish the 3 diagnostic groups (*normal*, *MCI*, and *dementia*), the Kruskal-Wallis test (with a Bonferroni adjustment for pairwise comparison) was performed on the following indicators: errors/targets ratio for TMT-A and TMT-B; ΔT for TMT-A, TMT-B, and Bells Test; and NIDmin for Bells Test. The trend of ΔT_n (TMT and Bells Test) and 1Bell (Bells Test) indicators according to the neuropsychological diagnosis was also studied, together with the scanning strategy during the Bells Test.

The significance level was set at 5% for all tests. Nonparametric statistical analyses were performed after verifying that the data were not normally distributed (Lilliefors test). The analyses were performed using RStudio version 1.0.143 (RStudio Inc).

Results

An example of the execution of the supervised neuropsychological test is provided in Multimedia Appendix 1 (the video presents the supervised execution of the 3 neuropsychological tests by an adult).

System Architecture Functional Testing

The correct identification of the semantic meaning of the users' utterances and the correct interaction flow were assessed on 3 subjects (2 women aged 25 and 26 years and 1 man aged 35 years).

The speech recognition functionality of the VA, evaluated over 200 tests, resulted in the detection of 92.5% (185/200) of the answers. The undetected answers can be ascribed to 2 main reasons: (1) excessive noise in the trials performed with added background noise and (2) wrong timing (eg, the user replied too soon, when the VA was not yet listening). A total of 91.9% (170/185) of detected answers were correctly recognized in the first attempt. In the remaining 8.1% (15/185) of cases, the VA automatically asked the user to repeat the answer. Only 1 repetition per answer was requested by the VA before it was able to correctly recognize the answer.

The workflow of the integrated system was tested in all possible execution branches, and no issues were encountered (100% (255/255) passed steps and 0% incomplete, blocked, and failed steps). During these runs, the system required the repetition of only 3 vocal inputs provided by the user. Given this result, the functional evaluation of the system architecture was considered successful. A demonstration of the interaction between the VA and TabletWebApp has also been proposed and tested during a series of live sessions [32,33].

Clinical Validation of the Digitized Tests

This study presents the results from 83 recruited patients. From the clinical neuropsychological visits, 27% (22/83) patients were categorized as *normal*, 59% (49/83) as *MCI*, and 15% as (12/83) *dementia* (Table 1). The age of the participants in the 3 diagnostic groups was not statistically different (P=.28). Among the patients selected by the neuropsychologist after the clinical visit, 6% (5/89) refused to take part in the study, whereas 1 patient who was recruited decided not to continue with the experimental trial because of an eye disease. All participants gave written informed consent for participation and authorization for the use of protected health information as reported in the ethical submission documentation.

Table 1. Characteristics of the participants (overall and divided into diagnosis groups).

Diagnosis groups	Normal	Mild cognitive impairment	Dementia	Total
Population, n	22	49	12	83
Age (years)				
Mean (SD)	76.2 (4.2)	78.0 (5.4)	78.6 (4.0)	77.6 (4.9)
Range	69-84	65-93	71-82	65-93
Mini-Mental State Examination score, mean (SD)	28.9 (1.1)	27.5 (2.2)	22.8 (1.9)	27.1 (2.7)
Education (years), mean (SD)	12.4 (4.4)	10.3 (4.6)	8.8 (5.1)	10.6 (4.9)
Sex, n (%)				
Female	16 (73)	23 (47)	8 (67)	47 (57)
Male	6 (27)	26 (53)	4 (33)	36 (43)

System Acceptance and Usability

The results obtained for system acceptance and usability are reported in Tables 2-4. The satisfaction questionnaire showed strong positive results. Indeed, most users (95% (21/22) normal, 91% (45/49) MCI, and 83% (10/12) dementia) were pleased with the overall experience. More specifically, the graphical interfaces of both the TMT (90.5% (20/22) normal, 83% (41/49) MCI, and 100% (12/12) dementia) and the Bells Test (81% (18/22) normal, 86% (42/49) MCI, and 83% (10/12) dementia) were considered simple and intuitive. In addition, the tutorial at the beginning of each test was found to be easy to understand by most users (100% (22/22) normal, 91% (45/49) MCI, and 100% (12/12) dementia). These data are particularly important if we consider that most users had limited familiarity with touch technology, as found from the questionnaires. Although some participants were able to use a computer (57% (13/22) normal, 37% (18/49) MCI, and 33% (4/12) dementia), few of them had used a smartphone (38% (8/22) normal, 17% (8/49) MCI, and 50% (6/12) dementia), and even fewer participants were acquainted with tablet technology (9.5% (2/22) normal, 6% (3/49) MCI, and 0% (0/12) dementia).

Table 2. System acceptance results from questions rated on a 3-point Likert scale.

General questions	Answers, n (%)			Diagnosis
	Agree	Neutral	Disagree	
I believe the test explanation and tutorial were clear				
	22 (100)	—[a]	—	Normal
	45 (91)	1 (3)	3 (6)	MCI[b]
	12 (100)	—	—	Dementia
I have found the graphical interface of the TMT[c] to be simple and intuitive				
	20 (90.5)	—	2 (9.5)	Normal
	41 (83)	5 (11)	3 (6)	MCI
	12 (100)	—	—	Dementia
I have found the graphical interface of the Bells Test to be simple and intuitive				
	18 (81)	—	4 (19)	Normal
	42 (86)	1 (3)	6 (11)	MCI
	10 (83)	—	2 (17)	Dementia
I am satisfied with the overall experience				
	21 (95)	1 (5)	—	Normal
	45 (91)	4 (9)	—	MCI
	10 (83)	—	2 (17)	Dementia

[a]Indicates 0%.

[b]MCI: mild cognitive impairment.

[c]TMT: Trail Making Test.

Table 3. System acceptance results from yes or no questions.

General questions	Answers, n (%)		Diagnosis
	Yes	No	
Have you ever used a smartphone?			
	8 (38)	14 (62)	Normal
	8 (17)	41 (83)	MCI[a]
	6 (50)	6 (50)	Dementia
Have you ever used a tablet?			
	2 (9.5)	20 (90.5)	Normal
	3 (6)	46 (94)	MCI
	—[b]	12 (100)	Dementia
Have you ever used a computer?			
	13 (57)	9 (43)	Normal
	18 (37)	31 (63)	MCI
	4 (33)	8 (67)	Dementia

[a]MCI: mild cognitive impairment.

[b]Indicates 0%.

Table 4. System usability results from the System Usability Scale questionnaires.

Diagnosis	SUS[a] score, mean (SD)	Participants with a positive score ≥68 (%), n (%)
Normal	82.0 (16)	19 (86)
Mild cognitive impairment	76.0 (17)	38 (77)
Dementia	77.5 (12)	10 (83)

[a]SUS: System Usability Scale.

In addition, for the SUS scores, positive results in terms of system usability emerged for all 3 groups: the average SUS score was 82 (SD 16) for normal, 76 (SD 17) for MCI, and 77.5 (SD 12) for dementia groups. Overall, 86% (19/22), 77% (38/49), and 83% (10/12) of the participants reported a positive SUS score (≥68) for the normal, MCI, and dementia groups, respectively.

System Validity

Digitized Test Versus Replay

For the TMT-A, the automatically identified error-TMT and target-TMT events by the software were compared with the ground truth that could be obtained by leveraging the replay functionality, reporting a sensitivity of 85.71%, specificity of 99.45%, and accuracy of 99.39%. For the TMT-B, a sensitivity of 72.16%, specificity of 98.39%, and accuracy of 96.84% was attained.

As for the Bells Test, the target indicator was strongly correlated with the one reported by the clinician through the replay (ρ=0.96; P<.001).

Digitized Test Versus Paper-Based Test

For both TMT-A and TMT-B, the completion time obtained in the digitized versions was strongly correlated with the completion time of the paper-based tests (TMT-A: ρ=0.68, P<.001; TMT-B: ρ=0.78, P<.001). Similarly, a significant strong correlation between the score of the digitized and the paper-based versions was found for TMT-BA (ρ=0.70; P<.001). On the other hand, the correlation between the targets in the digitized and paper-based versions of the Bells Test was significant but weak (ρ=0.39, P<.001). However, the Kruskal-Wallis test did not reveal any systematic difference between the 2 versions of the test as 45% (54/83) of the subjects achieved a better performance in the digitized test, whereas 44% (53/83) of them did better in the paper-based version, with a similar average number of targets identified in the 2 modalities of the test (digitized: mean 32.01 [SD 2.75]; paper-based: mean 31.80 [SD 3.74]).

Discriminative Power of the Neuropsychological Test Outcomes

For the TMT-A, the completion time obtained in the paper-based version was significantly different according to the diagnosis (X^2_2=22.6; P<.001), with values increasing as the diagnosis worsened. The posthoc analysis highlighted significant differences between all the groups (normal vs dementia: P<.001; normal vs MCI: P=.003; and MCI vs dementia: P=.02). As can be observed in Figure 4, the same significant trend was found for the digitized test (X^2_2=12.2; P=.002). However, the posthoc analysis showed a significant difference only between normal

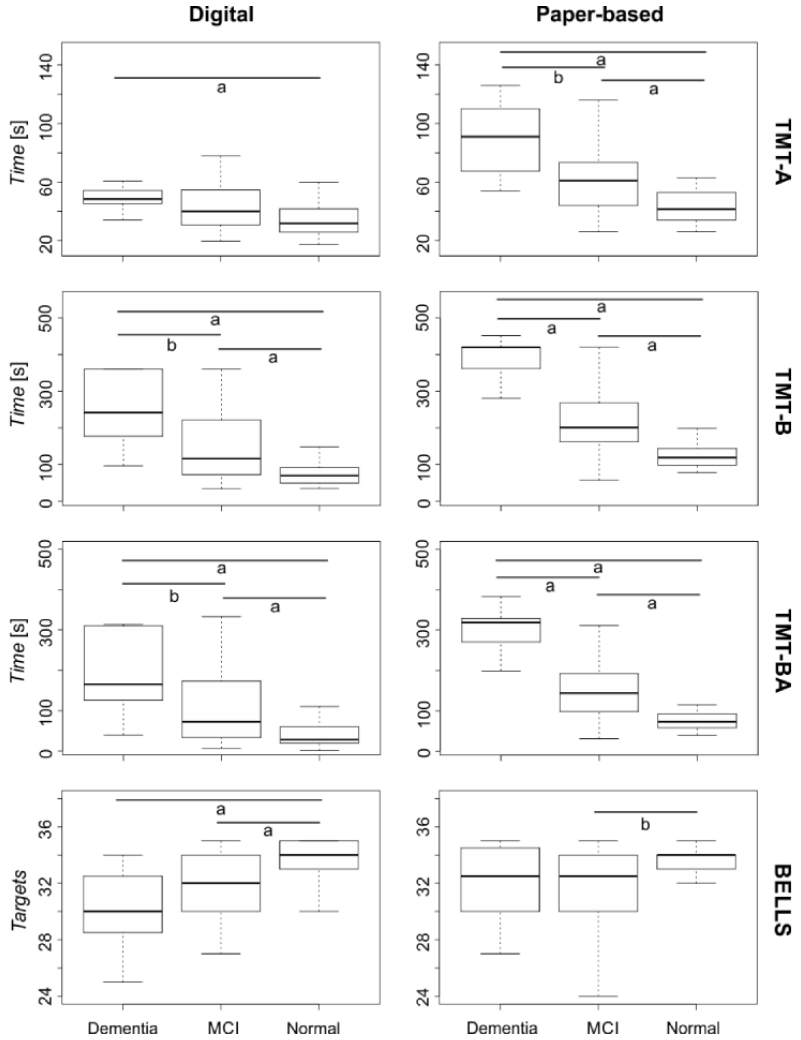
and dementia groups (P=.002) and an almost significant P value when comparing normal and MCI (P=.06) groups.

For the TMT-B, a significant effect of diagnosis on time emerged for both the paper-based (X^2_2=36.2; P<.001) and digitized (X^2_2=21.9; P<.001) tests, with values increasing as the diagnosis worsened. For both versions of the test, the posthoc analysis reported significant differences between all the groups (paper-based test: normal vs dementia P<.001, normal vs MCI P<.001, and MCI vs dementia P<.001; digitized test: normal vs dementia P<.001, normal vs MCI P=.004, and MCI vs dementia P=.02).

The TMT-BA showed similar results. Indeed, a significant effect of diagnosis emerged for both the paper-based (X^2_2=33.3; P<.001) and digitized (X^2_2=17.1; P<.001) versions, with more time required by patients with more severe symptoms. The posthoc analysis revealed significant differences between all the groups (paper-based test: normal vs dementia P<.001, normal vs MCI P<.001, and MCI vs dementia P<.001; digitized test: normal vs dementia P<.001, normal vs MCI P=.01, and MCI vs dementia P=.05).

Finally, for the Bells Test, the diagnosis had a significant effect on the targets of the paper-based (X^2_2=7.2; P=.03) and the digitized (X^2_2=18.9; P<.001) versions. However, as can be seen in Figure 4, only for the digitized version, we can observe a clear trend with fewer targets identified as the diagnosis worsened. Such observations were confirmed by the posthoc analysis. Although for the paper-based version, a difference emerged only between normal and MCI (P=.02), for the digitized test, the number of targets identified was significantly different between normal and dementia (P<.001) and normal and MCI (P<.001) groups.

Figure 4. Differences in test scores based on the participant's neuropsychological diagnosis. The boxplots represent, for the 3 neuropsychological diagnosis groups (normal, mild cognitive impairment, and dementia), the main outcome of the tests (time for Trail Making Test [TMT]-Part A, TMT-B, and TMT-BA and targets for Bells Test) as median and IQRs. Asterisks indicate significance from pairwise comparison (a indicates 1% significance and b indicates 5% significance).



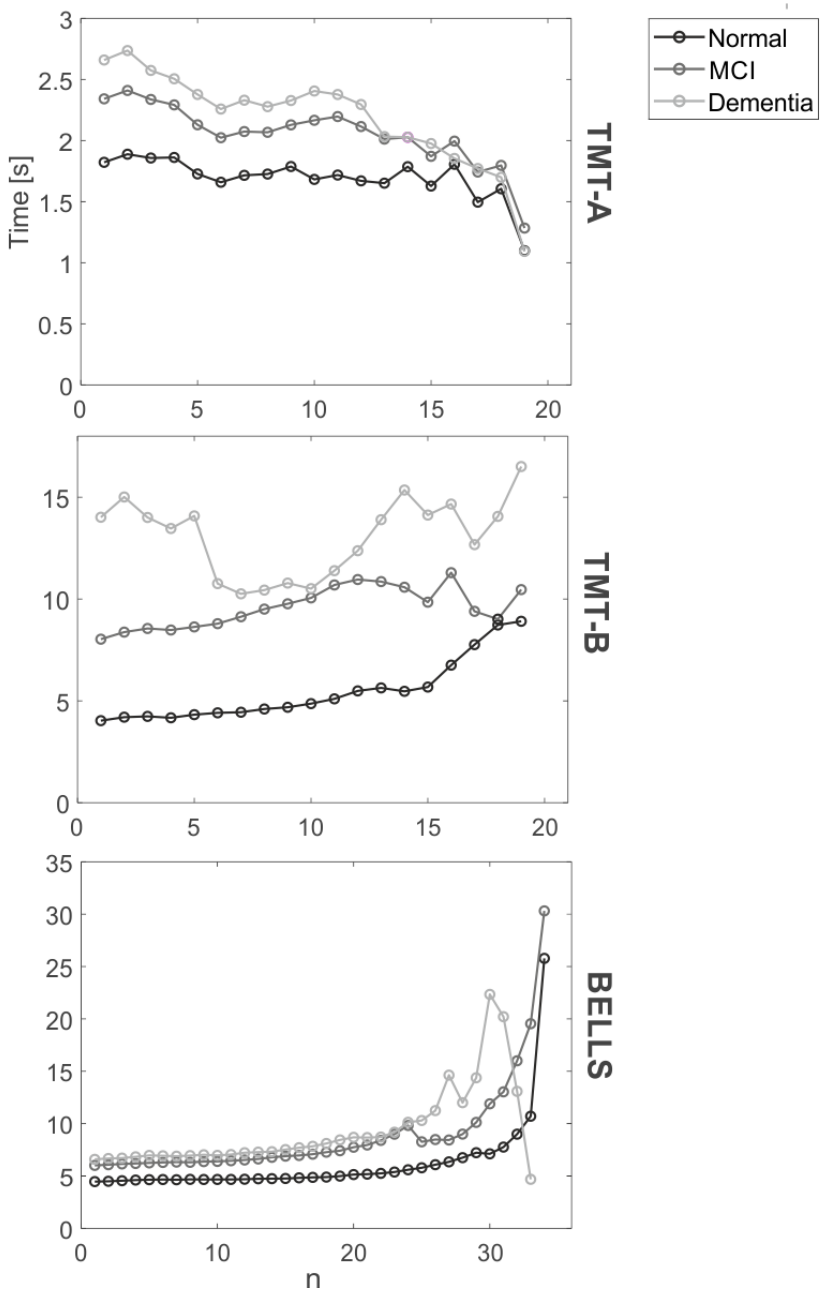
Discriminative Power of Novel Indicators Derived From the Digitized Tests

The errors/targets ratio was significantly affected by the diagnosis only for the TMT-B (normal: 0.05 [IQR 0.05]; MCI: 0 [IQR 0.1]; dementia: 0.15 [IQR 0.39]; X^2_2=8.6; P=.01). The posthoc analysis revealed that the errors/targets ratio of the dementia group significantly differed from both the MCI (P=.01) and the normal (P=.04) groups.

ΔT significantly changed according to the diagnosis for all the 3 tests (TMT-A: normal 1.67 [IQR 0.78] seconds, MCI 2.10 [IQR 1.24] seconds, dementia 2.55 [IQR 0.48] seconds, X^2_2=12.2, *P*=.002; TMT-B: normal 3.66 [IQR 2.14] seconds, MCI 6.10 [IQR 7.45] seconds, dementia 10.07 [IQR 9.14] seconds, X^2_2=21.1, *P*<.001; and Bells Test: normal 4.46 [IQR 1.11] seconds, MCI 5.88 [IQR 2.15] seconds, dementia 6.59 [IQR 2.94] seconds; X^2_2=13.5, *P*=.001). The posthoc analysis reported that, for the 3 tests, ΔT of the dementia group was significantly greater than that of the normal group (TMT-A: *P*=.002; TMT-B: *P*<.001; Bells Test: *P*=.004). ΔT for the TMT-B and Bells Test showed significant differences between the normal and MCI groups (TMT-B: *P*=.003 and Bells Test: *P*=.004). Only for the TMT-B, ΔT significantly differed between the MCI and dementia groups (*P*=.03).

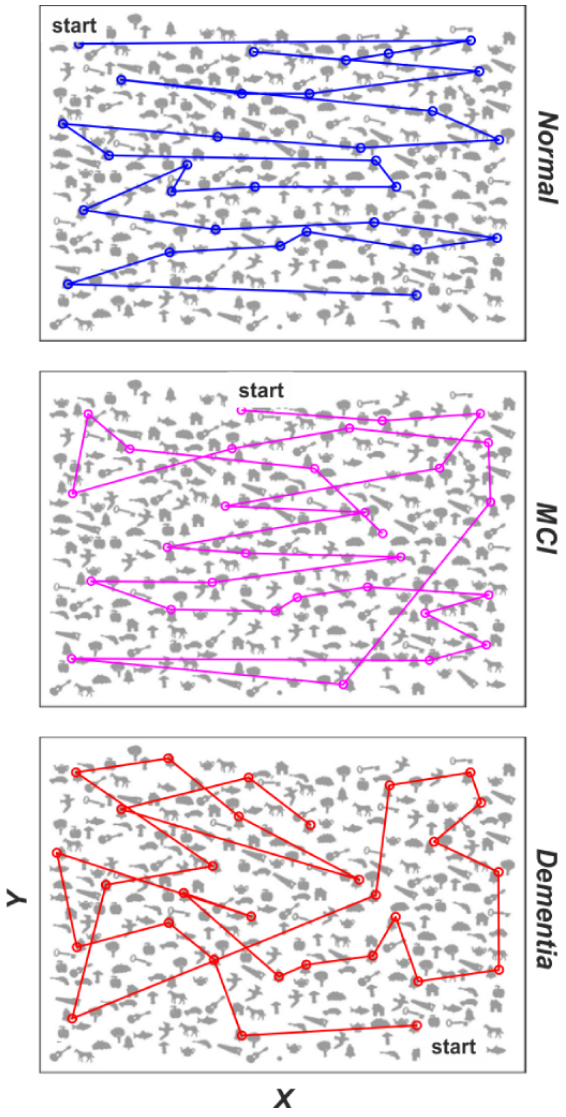
A qualitative analysis of ΔT_n was conducted and is presented in Figure 5. For all the 3 tests, the time between targets increases as the diagnosis worsens, as can be observed from the different offsets of the 3 lines. For TMT-A, the analysis revealed that ΔT_n decreased with test progression in all 3 patient groups. On the other hand, for all the groups, ΔT_n tended to increase toward the end of the TMT-B. As expected, for the Bells Test, the time between targets increased when identifying the final bells. In particular, a rapid increase can be observed after the 25th bell, especially in the dementia group. Note that the final drop of the dementia group curve is because of the fact that very few subjects were able to find more than 30 bells; therefore, the last points are obtained from the mean value over 1 or 2 patients and are not particularly representative of the entire dementia group.

Figure 5. From top to bottom, ΔT_n for Trail Making Test (TMT)-Part A, TMT-B, and Bells Test. For each test, the 3 lines represent the average ΔT_n over the diagnosis groups (black: normal; dark gray: mild cognitive impairment; and light gray: dementia).



Digitized testing is particularly suitable to gain insights into the subject's visuo-attentional scanning during the Bells Test execution. Figure 6 reports representative examples of visuo-attentional strategies that emerged during the digitized Bells Test for the 3 diagnostic groups. As can be observed, the regularity of visuo-attentional scanning decreases with MCI and even more with the onset of dementia. This behavior is reflected in the NIDmin indicator, which presents an increasing trend with the worsening of the diagnosis, although not statistically significant (normal: 1 [IQR 1.5], MCI: 2 [IQR 3], and dementia: 3 [IQR 3.5]).

Figure 6. Examples of scanning strategies during the digitized Bells Test for 3 subjects, one for each diagnosis group; from top to bottom: normal (NIDmin=0), mild cognitive impairment (NIDmin=2), and dementia (NIDmin=4). MCI: mild cognitive impairment; NIDmin: minimum number of inversions in the direction.



Another meaningful indicator that can be easily extracted and stored from the digitized Bells Test is related to the position of the first identified bell (1Bell), which is presented in Figure 7 (one panel for each diagnosis group). From Figure 7, it can be observed that although most of the subjects started the scanning strategy from the top-left corner (A1), this percentage decreases with the worsening of the diagnosis (normal: 74%, MCI: 66%, and dementia: 55%). In addition, the position of the first bell seems to be repeated among normal subjects, whereas an increased variability can be observed within the MCI group, for which the 1Bell indicator is spread over the canvas. In the dementia group, we noticed an additional reduction of the typical left-side start, although within-group variability decreases compared with MCI, as a significant percentage of patients with dementia (45% (5/12)) identified the first bell in the bottom-right corner (B3 and C3).

Figure 7. 1Bell for the 3 diagnosis groups (from top to bottom: normal, mild cognitive impairment, and dementia). For each group, the table represents the writing area of the tablet divided into 9 subareas and the percentage of subjects who started the scanning strategy in each subarea (1Bell). MCI: mild cognitive impairment.

Normal	A	B	C
1	74%	0%	16%
2	0%	0%	0%
3	11%	0%	0%

MCI	A	B	C
1	66%	9%	11%
2	2%	2%	2%
3	4%	4%	0%

Dementia	A	B	C
1	55%	0%	0%
2	0%	0%	0%
3	0%	18%	27%

Discussion

Novelty of Our System

Following the information and communication technology (ICT) revolution, computer-based versions of most classical neuropsychological tests have been developed in both research [12,13,15,34] and commercial [35] domains. However, the full potential of ICT technology is not captured by these approaches as the digitized tests must be either administered under the supervision of a clinical expert or self-administered without proper instructions and without including a professional in the loop. With the goal of providing a monitoring tool that offers a reliable administration of the tests, we developed a platform offering a supervised digitized version of 2 neuropsychological tests commonly used to assess dementia: the TMT, in its A and B parts, and the Bells Test. Our supervised tool aims to support screening for the early onset of age-related cognitive decline.

Although, to the best of our knowledge, no previous work ever implemented a digitized version of the Bells Test, previous literature developed a computerized TMT [12,13]. However, the novelty of our system originates from the intent of deploying the platform in an uncontrolled environment for which supervision on the modality of use is needed to guarantee the validity of the recorded data. A VA was designed to mimic the supervision offered by a neuropsychologist during the administration of the paper-based version of the tests: it delivered instructions to the user, recognized the user's answers, and produced the associated responses or interventions. The interaction between the VA and older adults is mediated through speech, as it is considered the most natural way of interacting with machines [27,36,37]. To this end, the dialogues were structured following the directions of clinical experts, and specific keywords and dictionaries were created to allow the VA to understand utterances with an equivalent semantic meaning. This way, the user can provide answers without constraints on the specific word, and the interaction is more robust and natural.

Together with web-based supervision, another key module of our system is the replay functionality, which was developed with the aim of allowing clinicians to remotely assess the test after execution. This functionality is crucial to increase the robustness and reliability of the system as it provides the clinician with the possibility of watching the test execution to better understand the obtained digitized score. Moreover, such functionality can be leveraged at clinical premises as the possibility of accessing test execution even at the end of the standard visit may potentially support the neuropsychologist in the diagnosis process, which could otherwise be hampered by the strict time schedule of the visits.

Principal Findings

As a first step, we successfully conducted a laboratory-based functional testing of the developed platform to test the robustness of the VA and its integration with the TabletWebApp.

A necessary second step, before the deployment of the developed computerized testing in an uncontrolled environment, is the study of user acceptance and test validity, which are known to be the 2 main barriers to achieving feasible computerized assessment [17]. The results obtained from the sample of 83 older individuals were very promising. Indeed, in terms of usability, most users reported a favorable score, independent of their diagnosis (86% (19/22) normal, 77% (38/49) MCI, and 83% (10/12) dementia. Concerning user acceptance, the questionnaire reported strongly positive results for all the 3 diagnosis groups, with most users satisfied with the overall experience and appreciating the clarity of the graphical interface. This result is even more appreciable if we consider the overall low familiarity with the technology of the recruited

subjects, especially with tablet devices. Our findings seem to confirm recent studies showing that although elders often portray themselves as digital illiterates, they show a surprising openness and ease of use when confronted with well-designed user-friendly technology [38]. In addition, it is worth pointing out that only 5 subjects refused to take part in the study, mostly because of exhaustion after the long neuropsychological visit, which took place right before the experimental session.

For system validity, we first leveraged the replay functionality to validate the software. For the Bells Test, the software was successfully validated, presenting a correlation coefficient between the 2 scores close to 1. In addition, for the TMT, satisfactory results were obtained in terms of specificity and accuracy, which were close to 100%. The sensitivity was slightly lower, mainly as some error-TMT events were misclassified as repetition-TMT when previously connected targets were hit again. Future work should focus on improving the sensitivity of the TMT software. Possible solutions may be the extension of the time-variant active area to all targets (and not restricted to the last one) and a postprocessing phase aimed at identifying possible misclassified error-TMT events.

We investigated both the concurrent and clinical validity of the computerized assessment compared with the traditional paper-and-pencil tests. For the TMT, the strong correlation between the 2 versions of the test suggests the potential of the digitized TMT to retain the same predictive power of the paper-and-pencil counterpart. This result confirms previous literature [13] for the digitized version of the TMT-B, whereas our results show a better outcome with regard to previous work on the TMT-A, for which only a moderate correlation was found between the digitized and paper-based versions. Concurrent validity is also supported by what emerged when investigating the effect of the neuropsychological diagnosis on the TMT score, with the time of completion increasing as the diagnosis worsened. Concerning the Bells Test, the weak correlation between the main outcome of the digitized and paper-based versions is somewhat surprising, especially given the fact that, contrary to the TMT, the layout of the digitized Bells Test was not modified with regard to the paper-based version, except for scaling down its layout size. Such differences in size are unlikely to explain the result as, owing to visual acuity, one would expect to identify a lower number of targets on a smaller support. However, in our case, there was no specific version of the test for which participants systematically achieved a better performance. A difference between the digitized and paper-based versions of the Bells Test also emerged when investigating the effect of the neuropsychological diagnosis on test scores. Although the result of the digitized test met our expectations, with the number of targets decreasing as dementia developed, the same trend was not found in the paper-based test, for which the number of bells identified by the dementia group was not statistically different from the normal group and was comparable with the MCI group. Thus, our results suggest a better discriminative power between mild and severe cognitive impairment for the digitized version of the Bells Test. Future work should investigate whether the reduced spatial contrast sensitivity that may characterize patients with dementia [39] could partially explain such a result.

One important advantage of computerized testing over traditional paper-based assessment is the possibility of quantifying additional information related to test execution strategies and not being restricted to the final test score. Such finer indicators may help clinicians gain insights into the patient's executive functions. In this framework, our work extracted novel indicators that can capture information related to the between-target time and the user's scanning strategy, which showed a decent discriminative power between diagnosis. In particular, the errors/targets ratio and ΔT indicators reported a very good between-diagnosis discriminant power, especially for the TMT-B; such a result is in line with previous literature on the paper-based TMT, which highlighted the effect of the diagnostic category only for TMT-B [40].

Depending on the test, different behaviors of ΔT_n indicator were observed. The TMT-A seemed to become easier toward the end of the execution (decrease in ΔT_n), and this trend appeared not to be affected by cognitive decline, as it was shared by all the 3 diagnostic groups. On the other hand, TMT-B test execution appeared to become more challenging for all the 3 groups when connecting the final targets, probably because of the mental fatigue induced by the test. In addition, for the Bells Test, the time between targets increased together with test progression, as expected.

In addition, the digitized version of the Bells Test allowed us to investigate the users' scanning strategy. Although the scanning order can potentially be studied for the paper-based test also, the procedure of writing down the bell identification order is time-demanding for the clinician, resulting in its nonapplicability in real practice. On the other hand, digitized testing allows extraction and visualization of the patient's scanning strategy in a convenient and easy way. As shown in Figure 7, the percentage of older adults starting the Bells Test scanning pattern from the top-left corner decreased with the severity of the diagnosis. Indeed, the normal group always started scanning from the sides, mostly from the left, as in the reading and writing processes. Instead, the MCI group showed the highest variability in the starting point of the scanning strategy. This variability decreased in the dementia group; however, almost half of the group identified the first bell in the bottom-right corner. An attempted explanation of what emerged is that, compared with healthy subjects, patients with MCI seem to be less attentive and present a mild early impairment of the planning strategy, reflecting slight executive function deficits [41]. In the dementia group, such deficits seem to be systematic, possibly revealing a visuo-attentional distortion.

Future Direction

Given the promising results obtained from the fine digital-specific indicators, future work should strengthen the collaboration between technicians and clinicians, within fruitful bidirectional translational research, with the aim of defining novel indicators that could make digitized tests even more useful. However, it is crucial to keep in mind that such computational tools should not and cannot replace human care providers or clinicians, but they should provide possible recommendations for seeking further professional help, thus

representing cost-effective solutions that are able to support the dementia screening programs of asymptomatic elders.

It is important to point out that, in this study, user acceptance and test validity were evaluated during an experimental protocol that did not include the VA. Indeed, digitized testing was performed at clinical premises under the supervision of a trained professional, without the guidance of the VA. Therefore, the reported results in terms of usability and concurrent validity do not apply to the full digitized testing platform; however, they represent a key and necessary step before the validation of the combined system. For this reason, we envisage that further usability and validity studies on the entire platform be performed in a home setting. To this end, the entire platform was deployed at the users' home within the European MoveCare Project [42], and further usability results will be available at the end of the pilot project.

Our system requires a reliable internet connection. Future work should explore solutions that are able to work without a stable internet connection to increase the accessibility of such tools and foster their use in rural areas. A possible solution should leverage open-source technologies for the development of VAs that also work offline, pushing the execution of core system functionalities to the user's side and introducing a local cache for data collection. In this way, we could loosen the persistent reliability of connectivity and allow the system to conduct all its functionalities under any network contingency. Indeed, functionalities running on the user's side can also work offline, and locally cached data can be resynchronized as soon as the communication channel is restored.

Finally, future work should investigate the possibility of juxtaposing traditional and strategy-related indicators in the creation of a predictive model that can boost the classification process between neuropsychological diagnoses.

Conclusions

Over the past decade, there has been a rapid evolution in the field of health screening methods to assist medical professionals to more accurately monitor older adults in relation to age-related conditions, such as dementia. With the goal of providing an effective tool to support the screening of cognitive decline, this work successfully designed and developed a supervised system for computerized neuropsychological assessment. The platform also includes a replay functionality designed to remotely inspect the user's performance after test completion. Test supervision and replay functionalities potentially allow the delivery of the system directly at the user's home or its use in clinical facilities without strict supervision to support the present medical services. Functional testing of the platform was successfully conducted. The digitized neuropsychological tests demonstrated very good concurrent validity, clinical validity, and a very high degree of user acceptance. The emerging positive results are necessary steps that pave the way toward the deployment of the system in clinical and nonclinical environments.

Acknowledgments

This work was supported by the European H2020 project MoveCare (Grant No. ICT-26-2016b–GA 732158. CA is now with IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy.

KD is now with the Department of Human Sciences for Education, *Riccardo Massa*, University of Milano-Bicocca, Milan, Italy, and the Department of Biomedical Sciences for Health, University of Milan, Milan, Italy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supervised execution of the 3 tablet-based neuropsychological tests by an adult.
[MP4 File (MP4 Video), 185313 KB - mhealth_v8i9e17963_app1.mp4]

References

1. World Health Organization. Dementia: A Public Health Priority. Geneva, Switzerland: WHO Press; 2012.
2. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999 Mar;56(3):303-308. [doi: 10.1001/archneur.56.3.303] [Medline: 10190820]
3. Chong MS, Sahadevan S. Preclinical Alzheimer's disease: diagnosis and prediction of progression. Lancet Neurol 2005 Sep;4(9):576-579. [doi: 10.1016/S1474-4422(05)70168-X] [Medline: 16109364]
4. Sirály E, Szabó A, Szita B, Kovács V, Fodor Z, Marosi C, et al. Monitoring the early signs of cognitive decline in elderly by computer games: an MRI study. PLoS One 2015;10(2):e0117918 [FREE Full text] [doi: 10.1371/journal.pone.0117918] [Medline: 25706380]
5. Budd D, Burns LZ, Guo Z, L'italien G, Lapuerta P. Impact of early intervention and disease modification in patients with predementia Alzheimer's disease: a Markov model simulation. Clinicoecon Outcomes Res 2011;3:189-195 [FREE Full text] [doi: 10.2147/CEOR.S22265] [Medline: 22046104]
6. Pasquier F. Early diagnosis of dementia: neuropsychology. J Neurol 1999 Jan;246(1):6-15. [doi: 10.1007/s004150050299] [Medline: 9987708]

7. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging Ment Health 2011 Nov;15(8):978-984. [doi: 10.1080/13607863.2011.596805] [Medline: 21777080]
8. Saxon R, Gray M, Oprescu F. Reducing geriatric outpatient waiting times: impact of an advanced health practitioner. Australas J Ageing 2018 Mar;37(1):48-53. [doi: 10.1111/ajag.12459] [Medline: 29044886]
9. Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open 2017 Feb 3;7(2):e011146 [FREE Full text] [doi: 10.1136/bmjopen-2016-011146] [Medline: 28159845]
10. American Psychological Association. Guidelines for Computer-Based Tests and Interpretations. New York, USA: The Association; 1986.
11. Canini M, Battista P, Della Rosa PA, Catricalà E, Salvatore C, Gilardi MC, et al. Computerized neuropsychological assessment in aging: testing efficacy and clinical ecology of different interfaces. Comput Math Methods Med 2014;2014:804723 [FREE Full text] [doi: 10.1155/2014/804723] [Medline: 25147578]
12. Dahmen J, Cook D, Fellows R, Schmitter-Edgecombe M. An analysis of a digital variant of the trail making test using machine learning techniques. Technol Health Care 2017;25(2):251-264 [FREE Full text] [doi: 10.3233/THC-161274] [Medline: 27886019]
13. Fellows RP, Dahmen J, Cook D, Schmitter-Edgecombe M. Multicomponent analysis of a digital trail making test. Clin Neuropsychol 2017 Jan;31(1):154-167 [FREE Full text] [doi: 10.1080/13854046.2016.1238510] [Medline: 27690752]
14. Lunardini F, Luperto M, Daniele K. Validity of Digital Trail Making Test and Bells Test in Elderlies. In: International Conference on Biomedical & Health Informatics. 2019 Presented at: BHI'19; May 19-22, 2019; Chicago, IL, USA. [doi: 10.1109/bhi.2019.8834513]
15. Maruff P, Thomas E, Cysique L, Brew B, Collie A, Snyder P, et al. Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. Arch Clin Neuropsychol 2009 Mar;24(2):165-178. [doi: 10.1093/arclin/acp010] [Medline: 19395350]
16. Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, et al. Validity of a novel computerized cognitive battery for mild cognitive impairment. BMC Geriatr 2003 Nov 2;3:4 [FREE Full text] [doi: 10.1186/1471-2318-3-4] [Medline: 14594456]
17. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):13-28. [doi: 10.1177/1533317514522852] [Medline: 24526761]
18. Gualtieri C, Johnson L. Reliability and validity of a computerized neurocognitive test battery, CNS vital signs. Arch Clin Neuropsychol 2006 Oct;21(7):623-643 [FREE Full text] [doi: 10.1016/j.acn.2006.05.007] [Medline: 17014981]
19. Sahakian B, Owen A. Computerized assessment in neuropsychiatry using CANTAB: discussion paper. J R Soc Med 2020:399-402 [FREE Full text]
20. Werner P, Korczyn AD. Willingness to use computerized systems for the diagnosis of dementia: testing a theoretical model in an Israeli sample. Alzheimer Dis Assoc Disord 2012;26(2):171-178. [doi: 10.1097/WAD.0b013e318222323e] [Medline: 21666431]
21. Dwolatzky T, Dimant L, Simon ES, Doniger GM. Validity of a short computerized assessment battery for moderate cognitive impairment and dementia. Int Psychogeriatr 2010 Aug;22(5):795-803. [doi: 10.1017/S1041610210000621] [Medline: 20519066]
22. Fillit HM, Simon ES, Doniger GM, Cummings JL. Practicality of a computerized system for cognitive assessment in the elderly. Alzheimers Dement 2008 Jan;4(1):14-21. [doi: 10.1016/j.jalz.2007.09.008] [Medline: 18631946]
23. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. Percept Mot Skills 2016 Aug 31;8(3):271-276. [doi: 10.2466/pms.1958.8.3.271]
24. Gauthier L, Dehaut F, Joanette Y. The bells test: a quantitative and qualitative test for visual neglect. Int J Clin Neuropsychol 1989;11(2):49-52. [doi: 10.1037/t28075-000]
25. Dorst K. The core of 'design thinking' and its application. Des Stud 2011 Nov;32(6):521-532. [doi: 10.1016/j.destud.2011.07.006]
26. Hashimoto R, Meguro K, Lee E, Kasai M, Ishii H, Yamaguchi S. Effect of age and education on the trail making test and determination of normative data for Japanese elderly people: the Tajiri project. Psychiatry Clin Neurosci 2006 Aug;60(4):422-428 [FREE Full text] [doi: 10.1111/j.1440-1819.2006.01526.x] [Medline: 16884442]
27. Schlögl S, Chollet G, Garschall M, Tscheligi M, Legouverneur G. Exploring Voice User Interfaces for Seniors. In: Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments. 2013 Presented at: PETRA'13; May 29-31, 2013; Island of Rhodes, Greece. [doi: 10.1145/2504335.2504391]
28. Giovagnoli AR, Del Pesce M, Mascheroni S, Simoncelli M, Laiacona M, Capitani E. Trail making test: normative values from 287 normal adult controls. Ital J Neurol Sci 1996 Aug;17(4):305-309. [doi: 10.1007/BF01997792] [Medline: 8915764]
29. Vallar G, Rusconi M, Fontana S, Musicco M. Tre test di esplorazione visuo-spaziale: taratura su 212 soggetti normali. Archivio di Psicologia 1994;55(4):827-841.
30. Brooke J. SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry. New York, USA: Taylor & Francis; 1996:189-194.

31. Huson LW. Performance of some correlation coefficients when applied to zero-clustered data. J Mod App Stat Meth 2007 Nov 1;6(2):530-536. [doi: 10.22237/jmasm/1193890560]
32. Luperto M, Romeo M, Lunardini F. Digitalized Cognitive Assessment Mediated by a Virtual Caregiver. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. 2018 Presented at: IJCAI'18; July 13-19, 2018; Stockholm. [doi: 10.24963/ijcai.2018/856]
33. Luperto M, Romeo M, Lunardini F. Evaluating the Acceptability of Assistive Robots for Early Detection of Mild Cognitive Impairment. In: International Conference on Intelligent Robots and Systems. 2019 Presented at: IROS'19; November 3-8, 2019; Macau, China. [doi: 10.1109/iros40897.2019.8968234]
34. Chan JY, Kwong JS, Wong A, Kwok TC, Tsoi KK. Comparison of computerized and paper-and-pencil memory tests in detection of mild cognitive impairment and dementia: a systematic review and meta-analysis of diagnostic studies. J Am Med Dir Assoc 2018 Sep;19(9):748-56.e5. [doi: 10.1016/j.jamda.2018.05.010] [Medline: 29921507]
35. Giunti Testing. Giunti Psychometrics. URL: http://www.giuntipsy.it/giunti-testing/ [accessed 2020-01-15]
36. Deng E, Mutlu B, Mataric M. Embodiment in Socially Interactive Robots. New York, USA: Now Publishers; 2019:251-356.
37. Fasola J, Mataric M. Evaluation of a Spatial Language Interpretation Framework for Natural Human-Robot Interaction With Older Adults. In: 24th IEEE International Symposium on Robot and Human Interactive Communication. 2015 Presented at: RO-MAN'15; August 31-September 4, 2015; Kobe, Japan. [doi: 10.1109/roman.2015.7333611]
38. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 4;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
39. Yan JH, Rountree S, Massman P, Doody RS, Li H. Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. J Psychiatr Res 2008 Oct;42(14):1203-1212. [doi: 10.1016/j.jpsychires.2008.01.006] [Medline: 18280503]
40. Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail making test errors in normal aging, mild cognitive impairment, and dementia. Arch Clin Neuropsychol 2008 Mar;23(2):129-137 [FREE Full text] [doi: 10.1016/j.acn.2007.11.005] [Medline: 18178372]
41. Brandt J, Aretouli E, Neijstrom E, Samek J, Manning K, Albert MS, et al. Selectivity of executive function deficits in mild cognitive impairment. Neuropsychology 2009 Sep;23(5):607-618 [FREE Full text] [doi: 10.1037/a0015851] [Medline: 19702414]
42. Lunardini F, Luperto M, Romeo M. The MOVECARE Project: Home-Based Monitoring of Frailty. In: International Conference on Biomedical & Health Informatics. 2019 Presented at: BHI'19; May 19-22, 2019; Chicago, IL, USA. [doi: 10.1109/bhi.2019.8834482]

Abbreviations

API: application programming interface
CAN-TAB: Cambridge Neuropsychological Test Automated Battery
CNSVS: CNS Vital Signs
GUI: graphical user interface
IRCCS: Istituti di Ricovero e Cura a Carattere Scientifico
MCI: mild cognitive impairment
MQTT: Message Queue Telemetry Transport
NID: number of inversions in the direction
NIDmin: minimum number of inversions in the direction
SUS: System Usability Scale
TMT: Trail Making Test
TMT-A: Trail Making Test-Part A
TMT-B: Trail Making Test-Part B
TMT-BA: Trail Making Test-Part B minus A
VA: vocal assistant

Edited by G Eysenbach; submitted 24.01.20; peer-reviewed by A Seelye, J Martínez-Miranda; comments to author 10.04.20; revised version received 12.05.20; accepted 03.06.20; published 21.09.20.

Please cite as:
Lunardini F, Luperto M, Romeo M, Basilico N, Daniele K, Azzolino D, Damanti S, Abbate C, Mari D, Cesari M, Borghese NA, Ferrante S
Supervised Digital Neuropsychological Tests for Cognitive Decline in Older Adults: Usability and Clinical Validity Study
JMIR Mhealth Uhealth 2020;8(9):e17963
URL: http://mhealth.jmir.org/2020/9/e17963/
doi:10.2196/17963
PMID:32955442

Original Paper

Efficiency of Text Message Contact on Medical Safety in Outpatient Surgery: Retrospective Study

Jeremy Peuchot[1], MD; Etienne Allard[2], MD; Bertrand Dureuil[1], MD, PhD; Benoit Veber[1], MD, PhD; Vincent Compère[1,3,4], MD, PhD

[1]Department of Anesthesiology and Critical Care, Rouen University Hospital, Rouen, France

[2]Le Havre Hospital, Le Havre, France

[3]Day Surgery Unit, Rouen University Hospital, France

[4]Normandie University, Institut National de la Santé et de la Recherche Médicale Unité 982, Rouen, France

Corresponding Author:
Vincent Compère, MD, PhD
Department of Anesthesiology and Critical Care
Rouen University Hospital
1 rue de Germont
Rouen
France
Phone: 33 02 32 88 82 83
Email: Vincent.compere@chu-rouen.fr

Abstract

Background: Establishing pre- and postoperative contact with patients is part of successful medical management in outpatient surgery. In France, this is mostly done via telephone. Automated information with short message service (SMS) reminders might be an interesting alternative to increase the rate of compliance with preoperative instructions, but no study has shown the safety of this approach.

Objective: The objective of this study was to evaluate the impact of pre- and postoperative automated information with SMS reminders on medical safety in outpatient surgery.

Methods: We conducted a retrospective, single-center, nonrandomized, controlled study with a before-after design. All adult patients who had outpatient surgery between September 2016 and December 2017 in our university hospital center were included. Before April 2017, patients were contacted by telephone by an outpatient surgery nurse. After April 2017, patients were contacted by SMS reminder. All patients were contacted the day before and the day after surgery. Patients contacted by SMS reminder were also contacted on day 7 after surgery. The primary end point was the conversion rate to full-time hospitalization. Secondary end points were hospitalization causes (anesthetic, surgical, organizational) and hospitalization costs.

Results: A total of 4388 patients were included, 2160 before and 2228 after the introduction of SMS reminders. The conversion rate to full-time hospitalization was 34/4388 (0.77%) with a difference between SMS group (8/2228, 0.36%) and telephone group (26/2160, 1.20%). The cost of SMS reminders was estimated as half that of telephone calls.

Conclusions: In this work, we report a decrease in the rate of conversion to full-time hospitalization with the use of pre- and postoperative SMS reminders. This new approach could represent a safe and cost-effective method in an outpatient surgery setting.

(JMIR Mhealth Uhealth 2020;8(9):e14346) doi:10.2196/14346

KEYWORDS

outpatient surgery; short message service (SMS); patient information; organizational; cost; unanticipated admission; preoperative instructions

Introduction

Outpatient surgery represents a major challenge in the organization of care. The increasing number of outpatient surgeries highlights the need to ensure the highest level of safety for each patient. Establishing contact with patients before and after surgery is part of successful medical management in

outpatient surgery, and it is strongly recommended by several practice guidelines [1,2].

Until recently, outpatients in France were mainly contacted by telephone before their surgery [3]. Telephone contact is costly and time-consuming with no guarantee of actually reaching the patient. According to Gaucher et al [4], the rate of telephone call failure to contact outpatients was 20.82% (425/2041 patients), whereas everyone can receive a short message service (SMS) text message so long as their mobile telephone is switched on. The use of SMS to contact surgery outpatients seems an interesting alternative to the telephone and has already been studied in chronic diseases and adherence to long-term treatments.

Several studies have shown the benefit of using SMS reminders [5,6] in patients with high blood pressure to decrease systolic blood pressure compared with usual care at 12 months or improve medication adherence. Others have developed a bank of text messages for the prevention of recurrent cardiovascular events [7]. A bank of 137 mobile telephone text messages designed to support behavior change and decrease cardiovascular risk has been developed through a multistep iterative process. During the testing of those 137 text messages, 92% of participants found the messages easy to understand and 86% found the messages contained useful information.

SMS reminders and a smartphone app have been used successfully to monitor and reduce the alcohol consumption of military veterans from a median of 5.6 units per drinking day in the first week to 4.7 units by the last week during the 4 weeks of study [8]. In contrast, inconsistent results were found in suicide prevention. There was no significant effect on likelihood or severity of current suicidal ideation or likelihood of a suicide risk incident or on emergency department visits [9].

Patient safety is a major concern for all medical teams. In studies, safety is defined as the occurrence of perioperative complications, conversion to full-time hospitalization, or rehospitalization after outpatient surgery. The most frequently used indicator is the conversion rate to full-time hospitalization [10-13]. Previous studies have shown that the use of SMS reminders before outpatient surgery increased the rates of compliance with preoperative instructions [14] and reduced the number of cancellations in gastrointestinal endoscopy [15]. The objective of this study was to evaluate the impact of pre- and postoperative SMS messages on medical safety in outpatient surgery.

Methods

Study Design

This was a retrospective, single-center, nonrandomized, controlled study with a before-after design. All data were extracted retrospectively from the database about patients who were hospitalized in the adult outpatient surgery unit of a French university hospital center from September 2016 to December 2017. Two groups of patients were formed corresponding to two study periods: September 2016 to March 2017 and April 2017 to December 2017. Ethical approval for this study (No. E2019-11) was provided by the noninterventional research committee based at Rouen University Hospital, Rouen, France. The requirement for written informed consent was waived by the committee.

Patient Selection

Patients scheduled for orthopedic; ear, nose, and throat; oral; dental; gynecologic; vascular; and thoracic outpatient surgery were included. Patients were eligible if they were aged over 18 years and had a mobile telephone. Patients were excluded if they were pregnant, younger than age 18 years, or did not speak French [1].

Study Procedure

Before April 2017, the first group of patients (telephone group) was contacted by telephone the day before and the day after surgery by an outpatient surgery nurse (Figure 1). After April 2017, the second group of patients (SMS group) was contacted by text message the day before surgery and on day 1 and day 7 after surgery (Figure 1). The deployment of a new version of scheduling software allowed us to send an SMS reminder to all patients with scheduled outpatient surgery. As soon as the outpatient surgery was scheduled, a nurse was able to set up SMS reminders, which were automatically sent. Patients' mobile telephone numbers are stored in the Gestime software internally developed by our university hospital center and connected to Computerized Medical Records CDP (GIP Cpage).

Messaging started two days before surgery with an SMS reminder and the possibility of alerting the medical staff if the patient was unable to attend surgery. The patient was able to respond ALERT if there was a medical problem or if assistance was required. A response different to ALERT was categorized as an unexpected response. There was no obligation to respond to the SMS reminder (Figure 2). We noted the percentage of patients who received the message but did not have the ability to know if messages were read or not. The day before surgery, the patient received 3 messages informing them of the required time of arrival and location of the outpatient unit, fasting recommendations, and hygiene rules (Figure 2).

Figure 1. Course in outpatient surgery before and after short message service text reminders.

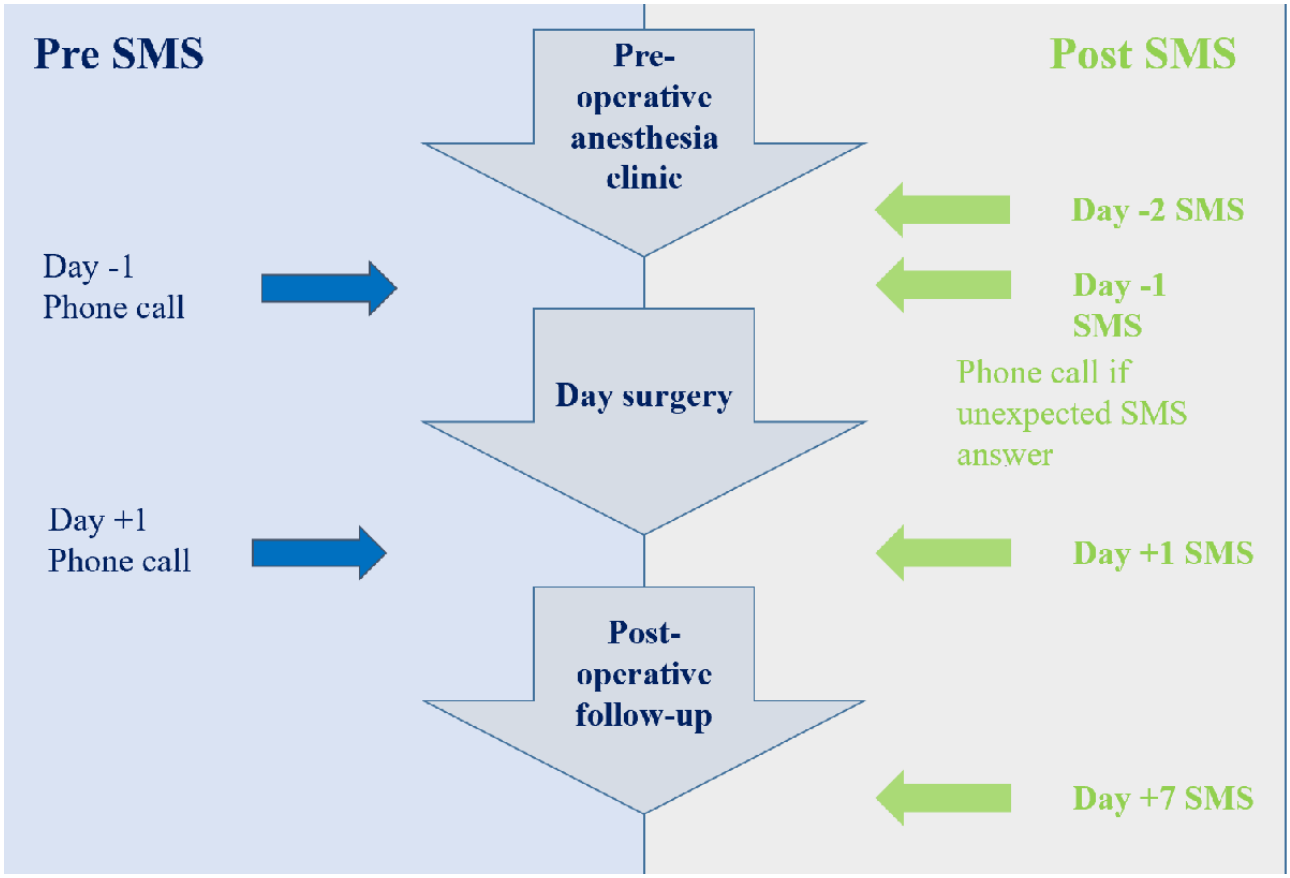
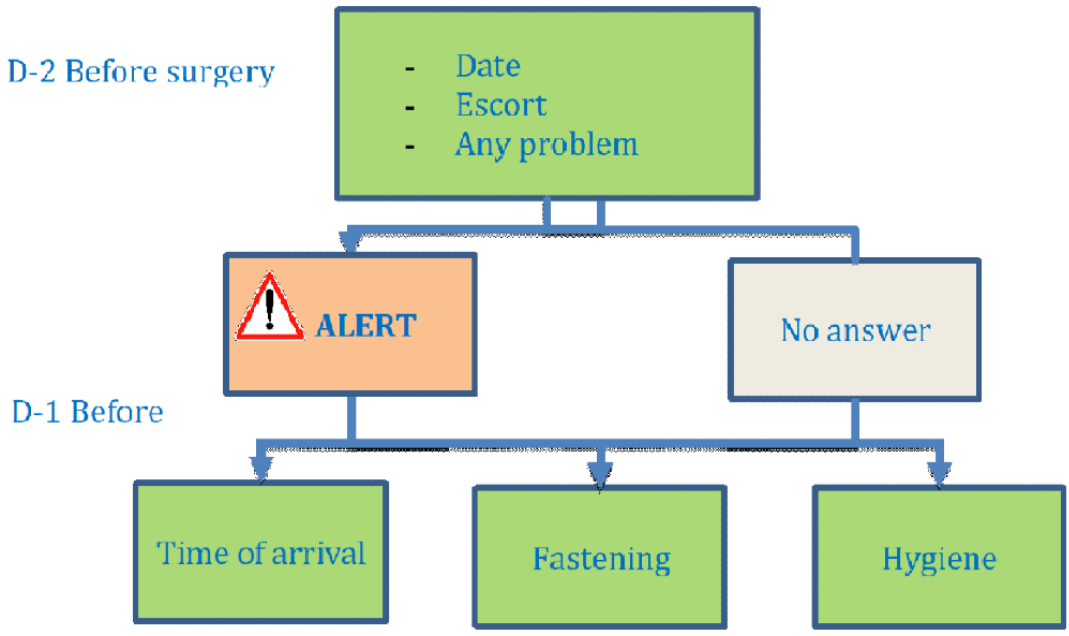
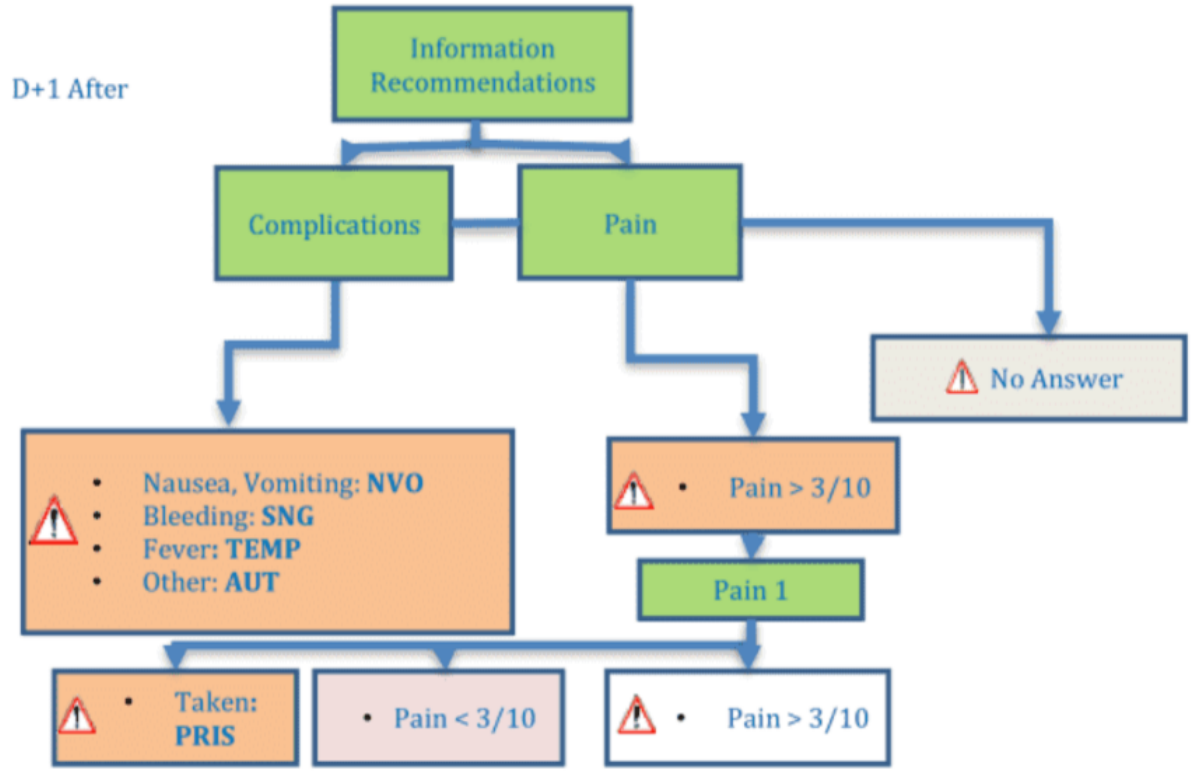


Figure 2. Short message service text pattern before surgery.



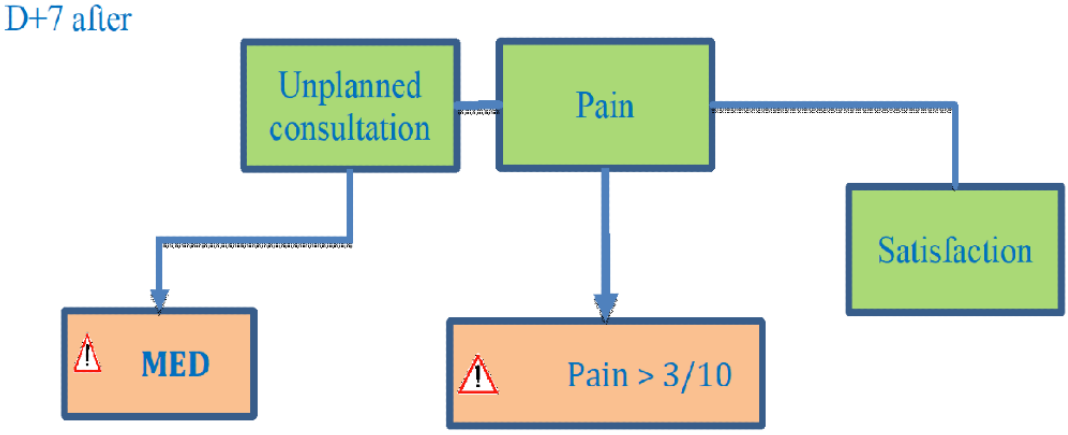
After surgery, patients received several SMS reminders with the possibility to answer. On day 1 after surgery, patients received an SMS reminder with recommendations on pain medication, a question about the onset of complications (nausea, vomiting, bleeding, fever, or other), and a pain evaluation on a numerical scale from 0 to 10 (Figure 3). If patients rated their pain more than 3/10, a reminder was sent to take their pain medication or immediately answer (TAKEN) if their pain medication had been taken. Patients were asked to send another pain evaluation 1 hour after taking their pain medication.

Figure 3. Short message service text pattern day 1 after surgery.



On day 7 after surgery, patients received one SMS reminder asking if they had had an unplanned consultation since their surgery and a second one asking them to evaluate their pain (Figure 4). Patients received one last message asking them to rate their global satisfaction with their outpatient care using a numerical scale from 0 to 10. Examples of SMS reminders are shown in Multimedia Appendix 1.

Figure 4. Short message service text pattern day 7 after surgery.



Data Collection

Data were extracted from the medical computerized database of our university hospital center. For hospitalized patients, anesthesia and surgery charts were reviewed. Perioperative data were collected.

- Preoperative: age, sex, BMI, comorbidities, allergies, smoking status, American Society of Anesthesiologists physical status classification (an assessment of the patient's preanesthesia medical comorbidities)
- Perioperative: type of surgery, type of anesthesia, length of surgery, time of end of surgery
- Postoperative: length of stay in postanesthesia care unit, pain evaluation using a 100-mm horizontal visual analog scale, need for complementary morphine titration, postoperative nausea and vomiting

SMS data were extracted from the computerized MemoQuest (Calmedica) database.

Outcome Variables

The primary end point was the conversion rate to full-time hospitalization. Secondary end points included cancellation rates on the day of surgery and a medicoeconomic study of SMS reminder use. In addition, causes of conversion to full-time

hospitalization (anesthetic: persisting pain, postoperative nausea and vomiting, sickness, aspiration syndrome; surgical: difficulty or length more than 2 hours; and organizational: entering or exiting operating room after 3:00 pm) and preoperative message (rate of ALERT replies, unexpected replies, undelivered SMS reminders) and postoperative pain (pain evaluation on day 1) data were collected. Last, data on pain evaluation and satisfaction and declarative rate of rehospitalization or unplanned medical consultation on day 7 postoperatively were collected, but we could not examine links between these different items.

Statistical Analysis

This study investigating the safety of SMS reminder contact before outpatient surgery did not require an a priori calculation of the number of subjects required. The values are presented as absolute and percentage for categorical variables and as mean and standard deviation for quantitative variables. The quantitative variables were compared using a Student *t* test (if the distribution was normal) or a Mann-Whitney *U* test (if the distribution was not normal). The categorical variables were analyzed using a Fisher or chi-square test. The significance threshold was set at .05. All statistics were produced using Prism 5.0 (GraphPad Software).

Results

Patient Characteristics

Between September 2016 and December 2017, 4388 patients were included in the study, 2160 in the telephone group and 2228 in the SMS group. Patient characteristics are presented in Table 1.

Table 1. Patient characteristics (n=4388).

Characteristic	Total (n=4388)	Telephone group (n=2160)	SMS[a] group (n=2228)	*P* value
Age in years, mean (SD)	43.2 (17.6)	43.0 (17.4)	43.3 (17.9)	.61
Sex, n (%)				
Women	3011 (68.62)	1466 (68.87)	1545 (69.34)	.67
Men	1377 (31.38)	694 (32.13)	683 (30.66)	.47
Length of surgery (min), mean (SD)	51.5 (32.6)	52.5 (30.6)	50.5 (31.6)	.04
Type of surgery, n (%)				
Thoracic	198 (4.51)	76 (3.52)	122 (5.48)	.002
Vascular	136 (3.10)	64 (2.96)	72 (3.23)	.67
Oral	296 (6.75)	168 (7.78)	128 (5.75)	.009
Dental	190 (4.33)	87 (4.03)	103 (4.62)	.37
Ear, nose, and throat	417 (9.50)	225 (10.42)	192 (8.62)	.04
Orthopedic	1578 (35.96)	817 (37.82)	761 (34.16)	.01
Gynecologic	1573 (35.85)	723 (33.47)	850 (38.15)	.001

[a]SMS: short message service.

Primary and Secondary Outcomes

The overall conversion rate to full-time hospitalization was 0.78% (34/4388); 1.20% (26/2160) in the telephone group versus 0.36% (8/2228) in the SMS group (*P*=.001). The conversion rates to full-time hospitalization and cause of hospitalization are shown in Table 2.

There was a significantly higher rate of cancellations declared on the day of surgery in the SMS group than in the telephone group: 3.66% (79/2160) versus 2.24% (50/2228); *P*=.02. There was no statistically significant difference in the demographic characteristics of patients converted to full-time hospitalization between groups (Table 3). Most of the surgeries were done under general anesthesia (25/34, 74%). Among all surgeries, orthopedic surgery accounted for 50% (17/34) and gynecologic surgery 44% (15/34).

In our study, 95.60% (2130/2228) of patients contacted by SMS reminder after their surgery actually received the reminder. For SMS reminder 1 sent on day 2 before surgery, 0.81% (18/2228) of patients replied ALERT and 6.24% (139/2228) sent an unexpected response. Most of the unexpected responses confirmed the patient's attendance at surgery.

For SMS reminder 2 sent on day 1 after surgery, 7.00% (156/2228) of patients rated their pain as more than 3/10. For SMS reminder 3 sent on day 7 after surgery, 2.60% (58/2228) of patients rated their pain as more than 3/10, and 6.69% (149/2228) of patients replied that they had seen their general practitioner or had been rehospitalized. A total of 90.98% (2027/2228) of patients gave a satisfaction score of 7/10 or more. The mean satisfaction score was 8.7 (SD 1.9).

Table 2. Conversion rates to full-time hospitalization between groups and according to the cause of hospitalization (n=4388).

Cause	Telephone group (n=2160)	SMS[a] group (n=2228)	*P* value
Total	26 (1.20)	8 (0.36)	.001
Anesthetic	15 (0.69)	4 (0.18)	.01
Surgical	7 (0.32)	3 (0.13)	.22
Organizational	4 (0.19)	1 (0.04)	.12

[a]SMS: short message service.

Table 3. Characteristics of patients converted to full-time hospitalization.

Variable	Telephone group (n=26)	SMS[a] group (n=8)	Total (n=34)
Type of anesthesia, n (%)			
General	20 (77)	5 (63)	25 (74)
General + regional	4 (15)	1 (13)	5 (15)
Regional	1 (4)	1 (13)	2 (6)
Spinal	1 (4)	1 (13)	2 (6)
Type of surgery, n (%)			
Orthopedic	12 (46)	5 (63)	17 (50)
Gynecologic	12 (46)	3 (38)	15 (44)
Oral	2 (8)	0 (0)	2 (6)
Length of surgery (min), mean (SD)	84.4 (60.7)	88.7 (58.1)	85.5 (58.6)
Length of stay in PACU[b] (min), mean (SD)	87.5 (37.1)	75 (58.6)	84.6 (50.5)
VAS[c], mean (SD)	3.3 (3.3)	1.7 (2.9)	2.9 (3.1)
Morphine titration, n (%)	4 (15)	2 (25)	6 (18)
Morphine dose (mg), mean (SD)	1.8 (4.2)	1.7 (3.2)	1.7 (3.7)
Prevention of PONV[d], n (%)	22 (85)	7 (88)	29 (85)
PONV, n (%)	24 (92)	8 (100)	32 (94)
Length of hospitalization in days, mean (SD)	3.0 (10.4)	2.5 (3.9)	2.9 (8.5)

[a]SMS: short message service.

[b]PACU: postanesthesia care unit.

[c]VAS: visual analog scale.

[d]PONV: postoperative nausea and vomiting.

Economic Aspect of Short Message Service Versus Telephone

Until recently, outpatients were mainly contacted by telephone by an outpatient surgery nurse before their surgery. However, telephone contact is time-consuming, representing 15 minutes per patient. In our study, there were approximately 3600 outpatients per year. At 15 minutes per patient, around 900 hours per year are spent contacting patients by telephone. Considering that the annual work time of a full-time nurse in France is 1607 hours [16], based on our findings, 56% of a nurse's time is spent contacting patients by telephone. Given that the annual cost of a full-time nurse is estimated at around €40,000 (US $45,018), the time spent contacting patients by telephone represents an overall cost of €22,400 (US $25,210) and an estimated cost of €6 (US $6.75) per patient. The use of an SMS reminder platform like MemoQuest (Calmedica), at an estimated cost of €3 (US $3.37) per patient [17], represents an annual cost of €10,800 (US $12,154) for 3600 patients. The cost of an SMS reminder was estimated as half that of a telephone call. The nurse time to check the SMS platform is negligible. One patient a day is called back after SMS reminder responses. The duration of the contact is much shorter (5 minutes) because the problem has already been focused by the SMS reminder.

Discussion

Principal Findings

In this work, we report a decrease in the rate of conversion to full-time hospitalization with the use of pre- and postoperative SMS reminders in an outpatient surgery setting. The rate of conversion to full-time hospitalization is one of the main criteria for monitoring safety in outpatient surgery [18]. We found a

lower rate of conversion (1.2%) compared with other studies. Cousin et al [19] in 2012 observed a rate of 3.1% in a retrospective study including 15,994 patients. Whippey et al [20], in a case-control study with 20,657 patients, observed a 2.7% conversion rate. Fortier et al [12], in a prospective study published in 1998 including 15,172 patients, reported a hospitalization rate of 1.4%.

Several hypotheses may underlie this decrease in the rate of conversion after the implementation of SMS reminders. First, the use of an SMS reminder before day surgery may be an efficient tool to confirm oral information given during a preanesthesia consultation. Garnier et al [14] showed that the use of SMS reminders significantly increased compliance with preoperative instructions, time of arrival, fasting rules, and hygiene rules. In our work, this decrease was associated with anesthetic causes, defined by persisting pain, postoperative nausea and vomiting, sickness, and aspiration syndrome. It is likely that an increase in compliance with preoperative instructions is associated with a decrease in intraoperative complications. Second, the use of SMS reminders may better identify subgroups of patients likely to cancel their surgery. This hypothesis is supported by the rate of cancellations observed on the day of surgery, which was higher in the SMS group compared with the telephone group (3.5% vs 2.3%). In a randomized controlled trial, Hashim et al [21] showed that the use of telephone reminders the day before an appointment significantly increased the rate of appointment cancellations, from 9.9% to 17% (P<.001). Moreover, our center has set up a waiting list of surgery outpatients to replace last-minute cancellation and thus avoid periods of inactivity in the operating room. Third, the use of SMS contact after surgery could improve patients' follow-up compared with nurse telephone calls. In our study, 95.6% of patients contacted by SMS reminder after their surgery actually received the reminder. This rate seems to be higher than that found in the literature for telephone contact. In the study by Hwa and Wren [22], 78% of all outpatients were successfully contacted by telephone. According to Gaucher et al [4], this rate was 79%. With the use of SMS reminders, patients were able to respond to the reminder to report a medical problem and were then contacted by telephone by an outpatient surgery nurse. Others have demonstrated the efficiency of SMS contact on medical safety. In a study on the use of SMS reminders to follow up on colorectal surgery patients enrolled in an enhanced recovery program, Carrier et al [23] showed early alerts of postoperative complications and no missed complications.

For these reasons, we consider that the use of SMS reminders represents a safe and cost-effective method to contact patients in an outpatient surgery setting. Few studies have determined the cost and effectiveness of reminders in an outpatient surgery setting. In a meta-analysis on appointment reminders, any type of reminder, whether by SMS or telephone, increased surgery attendance rates compared with no reminder. Although similar attendance rates were found between telephone and SMS reminders, the cost of an SMS reminder was more than 2 times less than that of a telephone call [24]. One recent systematic review highlighted the usefulness of appointment reminders even if further optimization is required [25]. The impact of reminders on the health care system is not well assessed at a time when the optimization of medical time, a limited and expensive resource, is much discussed. In our work, we estimated the cost of a telephone call to be around €6 (US $6.75) including nurse time. The use of SMS reminders represents an approach that could reduce costs by one-half. Others also evaluated the social cost [26] using parameters as nonuse or the misuse of personnel time, equipment, or ward capacity in a period of shortage of resources and a political will to restrain the costs of the health care services. In a recent practice survey of French outpatient surgery units, most preoperative reminders were done by telephone by nurses (58%), and email or SMS reminders were little used (1%). Postoperative follow-up was not systematically performed but when it was done it was by telephone and mainly by nurses (70%) [3]. The use of SMS reminders instead of telephone would allow nurses to spend more time taking care of their patients on the ward and less time making time-consuming telephone calls.

Strengths and Limitations

Our study has several limitations such as a lack of randomization. There are significant differences in patient characteristics between the groups. The 2-minute difference in the length of surgery, although statistically significant, has little overall impact. This small difference is probably related to the fact that there was more thoracic surgery than orthopedic and ear, nose, and throat surgeries. Another limitation of our study is its retrospective single-center design. However, the fact that it was conducted in an outpatient surgery unit of a university hospital center ensured the standardized organization of care teams. Our study has several strengths, including a relatively large number of patients. Also, the study period was short, which allowed us to observe the direct impact of the SMS reminders and limitations of other confounding factors.

Conclusion

In this work, we report a decrease in the rate of conversion to full-time hospitalization with the use of pre- and postoperative SMS reminders. This new approach could represent a safe and cost-effective method in an outpatient surgery setting.

Acknowledgments

The authors are grateful to Nikki Sabourin-Gibbs, Rouen University Hospital, for writing assistance and review of the manuscript in English. Support was provided solely from departmental sources.

Authors' Contributions

JP performed study conception and design, acquisition of data, statistical analysis, and analysis and interpretation of data and wrote the manuscript draft. EA performed patient recruitment and manuscript revision. BD performed patient recruitment and manuscript revision. BV performed patient recruitment and manuscript revision. VC performed study conception and design, study coordination, interpretation of data, and manuscript revision. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of short message service texts in French and their translations in English.
[PNG File , 464 KB - mhealth_v8i9e14346_app1.png]

References

1. Société Française d'Anesthésie et de Réanimation, Formalis R. Prise en charge anesthesique des patients en hospitalisation ambulatoire. Ann Françaises d'Anesthésie Réanimation 2010 Jan;29(1):67-72. [doi: 10.1016/j.annfar.2009.12.008]
2. Association of Anaesthetists of Great Britain and Ireland, British Association of Day Surgery. Day case and short stay surgery: 2. Anaesthesia 2011 May;66(5):417-434 [FREE Full text] [doi: 10.1111/j.1365-2044.2011.06651.x] [Medline: 21418041]
3. Beaussier M, Albaladejo P, Sciard D, Jouffroy L, Benhamou D, Ecoffey C, SFAR Committee of Ambulatory Anaesthesia. Operation and organisation of ambulatory surgery in France. Results of a nationwide survey: the OPERA study. Anaesth Crit Care Pain Med 2017 Dec;36(6):353-357. [doi: 10.1016/j.accpm.2017.07.003] [Medline: 28826981]
4. Gaucher S, Boutron I, Marchand-Maillet F, Baron G, Douard R, Béthoux JP, AMBUPROG Group Investigators. Assessment of a standardized pre-operative telephone checklist designed to avoid late cancellation of ambulatory surgery: the AMBUPROG multicenter randomized controlled trial. PLoS One 2016;11(2):e0147194 [FREE Full text] [doi: 10.1371/journal.pone.0147194] [Medline: 26829478]
5. Bobrow K, Farmer A, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-Text Adherence Support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
6. Wald DS, Bestwick JP, Raiman L, Brendell R, Wald NJ. Randomised trial of text messaging on adherence to cardiovascular preventive treatment (INTERACT trial). PLoS One 2014;9(12):e114268 [FREE Full text] [doi: 10.1371/journal.pone.0114268] [Medline: 25479285]
7. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett M, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
8. Leightley D, Puddephatt J, Jones N, Mahmoodi T, Chui Z, Field M, et al. A smartphone app and personalized text messaging framework (InDEx) to monitor and reduce alcohol use in ex-serving personnel: development and feasibility study. JMIR Mhealth Uhealth 2018 Sep 11;6(9):e10074 [FREE Full text] [doi: 10.2196/10074] [Medline: 30206054]
9. Comtois KA, Kerbrat AH, DeCou CR, Atkins DC, Majeres JJ, Baker JC, et al. Effect of augmenting standard care for military personnel with brief caring text messages for suicide prevention: a randomized clinical trial. JAMA Psychiatry 2019 May 01;76(5):474-483 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.4530] [Medline: 30758491]
10. Recommandations organisationnelles pour la chirurgie ambulatoire. Haute autorité de santé. 2013. URL: https://www.has-sante.fr/upload/docs/application/pdf/2013-05/guide_securite_du_patient_en_chirurgie_ambulatoire.pdf [accessed 2020-07-02]
11. Marshall SI, Chung F. Discharge criteria and complications after ambulatory surgery. Anesth Analg 1999 Mar;88(3):508-517. [doi: 10.1097/00000539-199903000-00008] [Medline: 10071996]
12. Fortier J, Chung F, Su J. Unanticipated admission after ambulatory surgery: a prospective study. Can J Anaesth 1998 Jul;45(7):612-619. [doi: 10.1007/BF03012088] [Medline: 9717590]
13. Fleisher LA, Pasternak LR, Lyles A. A novel index of elevated risk of inpatient hospital admission immediately following outpatient surgery. Arch Surg 2007 Mar;142(3):263-268. [doi: 10.1001/archsurg.142.3.263] [Medline: 17372051]
14. Garnier F, Sciard D, Marchand-Maillet F, Theissen A, Mohamed D, Alberti C, et al. Clinical interest and economic impact of preoperative SMS reminders before ambulatory surgery: a propensity score analysis. J Med Syst 2018 Jul 03;42(8):150. [doi: 10.1007/s10916-018-1000-z] [Medline: 29968118]
15. Deng X, Wang Y, Zhu T, Zhang W, Yin Y, Ye L. Short message service (SMS) can enhance compliance and reduce cancellations in a sedation gastrointestinal endoscopy center: a prospective randomized controlled trial. J Med Syst 2015 Jan;39(1):169. [doi: 10.1007/s10916-014-0169-z] [Medline: 25476268]

16. Décret n°2002-9 du 4 janvier 2002 relatif au temps de travail et à l'organisation du travail dans les établissements mentionnés à l'article 2 de la loi n° 86-33 du 9 janvier 1986 portant dispositions statutaires relatives à la fonction publique hospitalière. Legifrance. 2008. URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000005632044&dateTexte=20080403 [accessed 2020-06-02]
17. La start-up Calmedica équipe l'AP-HP d'un outil de suivi à domicile des patients. 2017. URL: http://www.ticsante.com/la-start-up-Calmedica-equipe-l-AP-HP-d-un-outil-de-suivi-a-domicile-des-patients-NS_3773.html [accessed 2020-06-02]
18. Delpont M. Note d'information n° dgos/pf2/2016/164. 2016 May 24. URL: http://circulaire.legifrance.gouv.fr/pdf/2016/06/cir_40989.pdf [accessed 2020-06-02]
19. Cousin A, Chatel C, Stéphanazzi J, Musellec H, Gentili M. [Analysis of output failures in an ambulatory surgery unit]. Ann Fr Anesth Reanim 2012 Sep;31(9):734-735. [doi: 10.1016/j.annfar.2012.06.006] [Medline: 22920329]
20. Whippey A, Kostandoff G, Paul J, Ma J, Thabane L, Ma HK. Predictors of unanticipated admission following ambulatory surgery: a retrospective case-control study. Can J Anaesth 2013 Jul;60(7):675-683. [doi: 10.1007/s12630-013-9935-5] [Medline: 23606232]
21. Hashim M, Franks P, Fiscella K. Effectiveness of telephone reminders in improving rate of appointments kept at an outpatient clinic: a randomized controlled trial. J Am Board Fam Pract 2001;14(3):193-196 [FREE Full text] [Medline: 11355051]
22. Hwa K, Wren SM. Telehealth follow-up in lieu of postoperative clinic visit for ambulatory surgery: results of a pilot program. JAMA Surg 2013 Sep;148(9):823-827. [doi: 10.1001/jamasurg.2013.2672] [Medline: 23842982]
23. Carrier G, Cotte E, Beyer-Berjot L, Faucheron JL, Joris J, Slim K, Groupe Francophone de Réhabilitation Améliorée après Chirurgie (GRACE). Post-discharge follow-up using text messaging within an enhanced recovery program after colorectal surgery. J Visc Surg 2016 Aug;153(4):249-252. [doi: 10.1016/j.jviscsurg.2016.05.016] [Medline: 27423211]
24. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013;12:CD007458. [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
25. McLean S, Booth A, Gee M, Salway S, Cobb M, Bhanbhro S, et al. Appointment reminder systems are effective but not optimal: results of a systematic review and evidence synthesis employing realist principles. Patient Prefer Adherence 2016;10:479-499 [FREE Full text] [doi: 10.2147/PPA.S93046] [Medline: 27110102]
26. Bech M. The economics of non-attendance and the expected effect of charging a fine on non-attendees. Health Policy 2005 Oct;74(2):181-191. [doi: 10.1016/j.healthpol.2005.01.001] [Medline: 16153478]

Abbreviations

SMS: short message service

Edited by G Eysenbach; submitted 11.04.19; peer-reviewed by L Lavorgna, T Timmers, G Deckard, D Leightley; comments to author 30.05.19; revised version received 24.10.19; accepted 14.05.20; published 10.09.20.

Please cite as:
Peuchot J, Allard E, Dureuil B, Veber B, Compère V
Efficiency of Text Message Contact on Medical Safety in Outpatient Surgery: Retrospective Study
JMIR Mhealth Uhealth 2020;8(9):e14346
URL: https://mhealth.jmir.org/2020/9/e14346
doi:10.2196/14346
PMID:32909948

Original Paper

Implementing Facilitated Access to a Text Messaging, Smoking Cessation Intervention Among Swedish Patients Having Elective Surgery: Qualitative Study of Patients' and Health Care Professionals' Perspectives

Kristin Thomas[1], PhD; Marcus Bendtsen[1], PhD; Catharina Linderoth[1], BSc; Preben Bendtsen[1], PhD

Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Kristin Thomas, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
58381
Linköping,
Sweden
Phone: 46 13282546
Email: kristin.thomas@liu.se

Abstract

Background: There is strong evidence that short-term smoking cessation before surgery can reduce postoperative morbidity. There are, however, several structural problems in health care systems concerning how to implement smoking cessation interventions in routine practice for preoperative patients.

Objective: This study aimed to analyze the implementation of a text messaging, smoking cessation intervention targeting patients having elective surgery. Implementation of facilitated access (ie, referral from practitioners) and the perceived usefulness among patients were investigated. Elective surgery is defined as scheduled, nonacute surgery.

Methods: A qualitative study was carried out at two medium-sized hospitals in the south of Sweden. The implementation of facilitated access was investigated during a 12-month period from April 2018 to April 2019. Facilitated access was conceptualized as specialists recommending the text messaging intervention to patients having elective surgery. Implementation was explored in terms of perceptions about the intervention and behaviors associated with implementation; that is, how patients used the intervention and how specialists behaved in facilitating usage among patients. Two focus groups with smoking cessation specialists and 10 individual interviews with patients were carried out. Qualitative content analysis was used to analyze the data.

Results: Two main categories were identified from the focus group data with smoking cessation specialists: *implementation approach* and *perceptions about the intervention*. The first category, *implementation approach*, referred to how specialists adapted their efforts to situational factors and to the needs and preferences of patients, and how building of trust with patients was prioritized. The second category, *perceptions about the intervention*, showed that specialists thought the content and structure of the text messaging intervention felt familiar and worked well as a complement to current practice. Two categories were identified from the patient interview data: *incorporating new means of support from health care* and *determinants of use*. The first category referred to how patients adopted and incorporated the intervention into their smoking cessation journey. Patients were receptive, shared the text messages with friends and family, humanized the text messages, and used the messages as a complement to other strategies to quit smoking. The second category, *determinants of use*, referred to aspects that influenced how and when patients used the intervention and included the following: timing of the intervention and text messages, motivation to change, and perceptions of the mobile phone medium.

Conclusions: Smoking cessation specialists adopted an active role in implementing the intervention by adapting their approach and fitting the intervention into existing routines. Patients showed strong motivation to change and openness to incorporate the intervention into their behavior change journey; however, the timing of the intervention and messages were important in optimizing the support. A text messaging, smoking cessation intervention can be a valuable and feasible way to reach smoking patients having elective surgery.

(JMIR Mhealth Uhealth 2020;8(9):e17563) doi:10.2196/17563

KEYWORDS

mHealth; mobile health; text messages; health care; smoking cessation; patients with elective surgery; implementation

Introduction

Smoking is responsible for more than 60 diseases and is the single-most important preventable factor for disease and premature mortality [1]. In persons between 45 and 64 years of age, the proportion of daily smokers is higher than for any other age group: 10% for women and 9% for men [2]. When including occasional smokers, the proportions are 16% for women and 13% for men. Although the proportion of smokers in Sweden is less than in many other countries, tobacco is associated with 9.6% of the total disease burden [3]. Thus, around 6400 people die every year in Sweden due to smoking.

The negative impact of smoking on outcomes following elective surgery is well established [4-6]. Several large studies have shown that the risk of cardiovascular, respiratory, and wound-healing complications and even death within 30 days of operation is greater for smokers than nonsmokers [7,8]. For instance, one study showed that smokers (11.3%) had a considerably increased risk for postoperative complications compared to nonsmokers (7.5%), especially an increased risk for pulmonary complications. In fact, smoking was found to be an independent risk factor exemplified by the fact that smokers needed intensive care and prolonged postoperative hospital stays to a greater extent than nonsmokers [9]. Research also show that complications can be avoided even with short-term perioperative smoking cessation [10,11]. Findings in a Cochrane review, based on indirect comparisons and evidence from two small trials, show that interventions beginning 4-8 weeks before surgery and including weekly counseling were most likely to have a significant impact on complications and on long-term smoking cessation [10].

However, more knowledge is needed on how to organize smoking cessation support within health care systems, especially for preoperative patients. Research has shown that various factors determine how and to what degree smoking cessation is implemented in routine care. For instance, motivation, knowledge, training, and confidence among health care professionals to deliver smoking cessation support have been shown to have an influence [12-14]. Beliefs and preconceptions among health care professionals that question the value or benefit of smoking cessation support have also been reported; for example, some health care professionals believe that interventions are time-consuming, ineffective, or intrusive [15]. Furthermore, organizational challenges, including lack of standardized pathways or referral routines across and between parts of the health care system, have also been quoted [12]. Indeed, organizational solutions as to how to best implement smoking cessation support for this patient group is unclear, for example, whether smoking cessation support optimally is delivered in hospitals, via specialist care, or in primary care. Although primary care could be a useful setting for smoking cessation support before surgery, challenges to achieve working referral routines have been reported, including long waiting times, resistance from staff in primary health care to offer support to patients from specialist care, and unclear communication between specialist care and primary health care [16,17].

These hurdles to implement smoking cessation support in health care present a need for new types of innovations [17]. One way forward could be to give smokers having elective surgery access to support through their mobile phones. Mobile phone–based interventions that use text messaging typically encompass a series of automated messages during a period of about 8-12 weeks. Messages aim to guide and support participants to plan and prepare for a quit attempt and then reinforce and support continued smoking cessation [18].

There is a substantial body of evidence demonstrating the effect of text messaging on smoking cessation among adolescents [19-21], university students [22], and adults [18,21,23,24]. Furthermore, mobile phone–based interventions including text messaging have shown to be one of the most cost-effective interventions for tobacco control and are endorsed by the World Health Organization [25]. Major advantages have been shown to be cost-effectiveness, reach, and flexibility, whereby they can be delivered at any time [26].

In our previous research, we have developed evidence-based interventions that use text messaging: the NEXit (Nicotine Exit) and NEXit Junior trials [22,27]. These interventions are based on behavior change theory and evidence-based practice guidelines [28,29]. This study aimed to analyze the implementation of a text messaging, smoking cessation intervention targeting patients having elective surgery. Specifically, implementation of facilitated access (ie, referral from practitioners) and the perceived usefulness of the intervention among patients were investigated. Elective surgery is defined as scheduled, nonacute surgery.

Methods

Design

In this qualitative study, we investigated the implementation of facilitated access and perceived usefulness of a text messaging, smoking cessation intervention among patients having elective surgery. Two focus groups with smoking cessation specialists and 10 individual interviews with patients were carried out. The implementation of facilitated access (ie, referral from practitioners) was investigated during a 12-month period from April 2018 to April 2019. Implementation was explored in terms of perceptions about the intervention and behaviors associated with implementation; that is, how patients used the intervention and how specialists behaved in facilitating usage among patients.

Setting

The study was carried out at two medium-sized hospitals in the south of Sweden. The smoking cessation offices were located at the hospitals and received patients by referral from surgeons

and walk-ins. Referral came primarily from general, orthopedic, heart, and gastrointestinal surgery departments as well as, to a smaller extent, from pulmonary medicine, rheumatology, and cardiac clinics. Practice routines typically include an initial session with a patient where motivation to quit is explored, followed by follow-up meetings and/or telephone calls depending on the preferences of the patient.

The three smoking cessation specialists were asked to facilitate implementation by inviting patients to sign up for the intervention during a face-to-face visit. Sign-up was done by patients sending a text message with a specific code to a dedicated telephone number. The text messaging, smoking cessation intervention is an evidence-based, 12-week, fully automated text messaging program that has been described elsewhere [30].

In total, 100 patients were approached, and 30 patients signed on to the intervention during the visit, 44 patients wanted to think about signing up at a later stage at home, whereas 26 patients were not interested in the intervention. After signing up to the intervention, the patients received a text message with a link to a baseline questionnaire; a total of 27 patients completed the questionnaire and were thereafter enrolled in the intervention.

Data Collection and Participants

Smoking Cessation Specialists

The smoking cessation specialists were registered nurses with training in smoking cessation counseling. The training included a 3-day course to become registered smoking cessation specialists as well as ongoing supervision. All specialists were also qualified in motivational interviewing. All three smoking cessation specialists were invited and took part in focus group interviews. Data collection was conducted at two time points: at 6 and 12 months after starting to implement the intervention among their patients. The focus group interviews lasted approximately 1 hour; they followed a semistructured format and included questions on the experience of implementing the intervention in routine practice.

CL and PB took part in the first focus group and CL and KT took part in the second focus group. Participants were all women between 45 and 54 years of age.

Patient Interviews

Telephone interviews were conducted with patients 3 months after the intervention. A semistructured interview guide was used that aimed to capture patients' experiences and use of the intervention.

A total of 14 patients were contacted via telephone and invited to take part in interviews; 4 patients declined to participate. A total of 10 interviews were carried out and they lasted between 10 and 40 minutes. Respondents were between 45 and 70 years of age; the sample was made up of 6 men (60%) and 4 women (40%). Telephone interviews were either scheduled later or conducted straight away depending on patients' preferences.

Data Analyses

All focus group discussions and individual interviews were audio-recorded and transcribed verbatim. The two datasets—patients and specialists—were analyzed separately. The analysis process, which followed conventional content analysis guidelines by Hsieh and Shannon [31], was used to analyze the data. Conventional content analysis is a structured process where data relevant to the study aim are coded and categorized. First, all transcripts were read through. Second, words and text that depicted areas relating to the study aim were identified (coding). In a parallel process, these identified codes were grouped (categorizing) based on similarity in content and their relation to each other. The content of each category was expanded by revisiting the data and comparing data across formed categories. Lastly, a comparison across categories was done to make sure that categories were defined and described in a way that maximized internal homogeneity and external heterogeneity. KT performed initial data analysis, which was then discussed by KT and PB.

Ethical Approval

The study has received ethical approval by the Regional Ethical Review Board of Linköping University, Sweden (2018/4-31).

Results

Focus Group Interviews With Specialists

Overview

Two main categories were identified in the data that described the implementation of the intervention in routine smoking cessation practice: implementation approach and perceptions about the intervention. Initial coding resulted in a number of codes. Examples of these codes are given in Table 1. For example, prioritizing trust referred to how specialists chose to behave, communicate, and give information to patients that enhanced patients' trust. For instance, specialists actively chose to communicate the intervention at the end of sessions when initial trust between patients and themselves had been established.

Table 1. List of categories, subcategories, and code examples for specialist data.

Category and subcategories	Code examples	Excerpts of quotes from raw data[a]
Implementation approach		
Adapting		
	Time	"Sometimes I need to help them, but it is difficult to have enough time."
	Patient qualities	"[Some patients are] thorough and want to follow the program exactly, exemplary, and want to be best in the class so to speak."
	Patient capacity	"We have some really old...if we are to talk about age, for old people it is really difficult [to enroll]."
Trust building		
	Relationship to patient	"Yes, I usually have the same method; I try to create a good relationship with them in the first meeting."
	Prioritizing trust	"You have to focus on what's important and get that relationship first that you don't do by pressuring them."
Perceptions of the intervention		
Familiarity		
	Compatible with existing routines	"I think it's been ok to include in the session, like it hasn't been odd, it's been easy."
	Recognizing content	"And it is so much that you recognize in the text messages...so much and similar to what we use in our practice, so that's good."
Complementary		
	Add-on	"No, I've seen it as a something additional to what I do normally."
	Introducing at the end	"I think, also, that I introduce it at the end like the typical session talk first and then at the end introduce it."

[a]Translation from Swedish to English for the purposes of this publication only.

Implementation Approaches

Overview

The category *implementation approach* refers to how specialists implemented the intervention and strategies that they used: adapting to specific situations and building trust with patients before introducing the intervention.

Adapting

Adapting refers to how specialists altered their behavior and approach to facilitate implementation through, for instance, how they introduced the intervention, talked about it, or assisted the patient in signing up. Adaptations were influenced by patient needs and capacities as well as situational factors. For example, the patients were perceived to have varied capacities and capabilities to sign up and engage in the intervention. One respondent described how one patient was not able to sign up for the intervention due to feeling overwhelmed by the technical requirements of completing the registration and, thus, needed additional support.

> *When I sat with my patient and together completed the registration with him, he thought "What's this?" and in the end I had to take the phone and, like, do it for him.*

Similarly, the specialists used strategies to invite patients to sign up for the intervention depending on the needs and preferences of the patients. For example, specialists could help patients sign up and complete baseline questionnaires if this would enable patients to start the intervention.

> *It, you start with the patient, you always have the patient in mind, so there's lots of different needs for whoever sits there.*

Furthermore, adaptations were made due to situational factors such as time constraints. Respondents expressed that it was sometimes challenging to have enough time to introduce and sign up patients within one session. Adaptations were made regarding how they spoke about the intervention or delayed introducing the intervention until later sessions to make sure they would have enough time to engage patients in registration. Finally, one respondent expressed that when meeting patients that were "good student"–type people, it was easier to introduce the intervention and that this group of patients needed less guidance.

> *Yes, I think about the lonely people, it becomes a friend in a way, that somebody sends you text messages. And those who are really, some are very thoughtful and careful and want to do the right thing always.*

Trust Building

Respondents expressed that they tried to build trust with patients first and then, often at the end of the session, introduce the intervention. Building trust entailed creating a trusting relationship with the patient by listening to their needs and

pacing how much, and what kind of, information they shared with the patient. Respondents expressed that it was useful to invest in this relationship first and then introduce the intervention.

> *Yes, you can say that I have a way, you try to create a relationship with the patient at the first meeting and then finish with "Oh, we have this study," and then I tell a bit more about that.*

Perceptions About the Intervention

Overview

The category *perceptions about the intervention* refers to how specialists perceived the intervention: whether its content and structure were familiar, and whether the intervention was a complement to current practice.

Familiarity

In general, the respondents were positive toward the intervention and voiced that the concept, content, and structure of the intervention was in line with techniques that they already used in patient practice. Specialists expressed that that they experienced the intervention as familiar, for instance, that its content was straightforward, logical, and compatible with smoking cessation strategies that they already employed among patients. The respondents further described the intervention as feasible and easy to implement. In addition, respondents spoke about mobile phone–based interventions in general and expressed an interest and ease to incorporate these types of tools in practice.

> *And there's much, you recognize a lot that's written in the messages. We got them all on paper before, it is, you recognize, a lot that I use in my work, so I think it is good.*

Complementary

The intervention was perceived as a complement to be offered to patients alongside existing support and tools, such as follow-up phone calls or medication. The intervention was described as something that was introduced to most patients at the end of the routine smoking cessation session. Respondents perceived the intervention as a complement to what they already did, something parallel that did not necessarily affect established routines but a tool that was feasible to use.

> *No, but I have seen this as a complement to what I do, so doesn't disturb what I do, I haven't seen it in that way.*
>
> *It's been an add-on to what we already offer, now we can offer this as well: "Would this be something for you?"*
>
> *It feels like it's not what you want to start with; you have to focus on the main issue and create the relationship, and you can't do that if you pressure the patient.*

Furthermore, the intervention was described as a valuable alternative tool when there were limited resources, particularly for patients with complex needs.

> *Sure, we have some quite intensive patients that need more regular contact and follow-up visits. And then it feels nicer if they have accepted this, to know that, yes, they get something; meanwhile, if I can't book them in next week when I was supposed to, then I know that they will get some sort of support.*

Individual Interviews With Patients

Overview

Two main categories were identified in the data: incorporating new means of support from health care and determinants of use. See Table 2 for examples of these codes.

Table 2. List of categories, subcategories, and code examples for patient and specialist data.

Categories and subcategories	Code examples	Excerpts of quotes from raw data[a]
Incorporating new means of support from health care		
Being receptive		
	Curiosity	"I was waiting on the text messages every day, like so like you looked and wondered what have they written today."
	Openness	"So it was like a bit, ok, I've got [a message] from them again [laughs], so it has been really good I think."
Sharing		
	Impact beyond user	"I have sent screenshots to a friend when it has been good stuff."
	Communicating	"And we said, my wife always asked, 'Is it the smoking cessation people?' Yes, and now they write this and that."
Humanizing of text messages		
	Expectations	"To start again when they have put so much effort in, I felt it would be to let somebody down if I started again."
	Presence	"It feels like there is somebody sitting and sending the messages."
Text messages as a complement		
	Alternative strategies	"So I could do it at my pace and then, and then my wife said it doesn't matter if it takes 10 minutes or half hour, 'Let's go for a walk.'"
	Combination	"I think it was the combination, 'cause I don't think text messages alone would do it for me."
Determinants of use		
Timing		
	Relevance	"Yes I think then that...the text messages are really good when you're in the middle of it before surgery, then you get triggered."
	Timing	"But as I had already quit, then it was a bit stupid, this about preparing for quitting in two weeks, then I had already quit."
Motivation to change		
	Decisiveness	"Yes, I can see now, I have thought that there is no point or meaning to start again."
	Action	"And then I walked down to her at the unit straight away from meeting with the surgeon to speak with her."
Perceptions of the mobile phone medium		
	Reached	"It was a bit personal, as it came from a text message rather than an app."
	Limited interaction	"I didn't go on the links...you got to write why you quit...but I read all the text messages."

[a]Translation from Swedish to English for the purposes of this publication only.

Incorporating New Means of Support From Health Care

Overview

This category referred to how patients adopted and incorporated the intervention into their smoking cessation journey. Incorporating new means of support from health care encompassed the following: being receptive, sharing the text messages with friends and family, humanizing of text messages, and using the text messages as a complement to other strategies to quit smoking.

Being Receptive

Being receptive refers to having an openness to learn, to assimilate information, and to try techniques suggested in the text messages. Learning is talked about in terms of gaining knowledge about new things, such as how to manage abstinence but also about being reminded of things you already know; for example, the health risk of smoking. Respondents describe an openness to this learning process and that it was useful in their behavior change.

In addition, being receptive was illustrated by curiosity and interest among respondents regarding the content of the

messages. This can further be exemplified by a respondent that described that they waited for the text message to arrive in the morning, being eager to know what it would say. The data further showed that users read all the text messages, either when they arrived or later the same day if it was more convenient. Even respondents that expressed irritation from text messages described that they still read all the text messages and reflected on their content.

> *It is difficult to say. Now I waited for this...every day...like, you looked, "What did they say today?"*

Sharing

Part of processing the content of the intervention entailed sharing it with significant others. Respondents expressed that they read the text messages and shared what they learned or discrete messages with friends and family. For example, one respondent described that they, together with their partner, each day read the messages and reflected on their content. In this way, using the intervention was a shared activity, similar to the actual smoking cessation journey where family members would be described as an important influencer. Other respondents described that they forwarded messages to friends they knew smoked and who could benefit from the content.

> *I have a friend that also wants to quit, so I, like, sent the messages to her, the good ones, so I forwarded those to her, like.*

Humanizing of Text Messages

Respondents spoke about the text messages as personal and compatible with human support. Respondents described that receiving text messages made them feel less lonely, safer, and as if somebody cared about them. For example, one respondent perceived the text messages as a presence, as somebody who cheered them on and encouraged them to keep being smoke-free. Another respondent highlighted nonsmoking norms in society and how smokers are often a minority and that the text messages made him feel less abandoned.

> *Yes, I think so. Because I think that many people feel like this, that they feel lonely in all of this, like when they do it.*
>
> *No, I don't know, but I was really grateful that I got support and that somebody cared and that I wasn't alone with this to struggle with.*

Furthermore, there were similarities between the expectations that respondents placed on the text messages and people in their social environment. For example, one respondent compared the text messages to a nagging relative that wanted them to quit smoking. Similar to expectations in their social environment, this respondent expressed that they did not appreciate nagging but preferred encouraging messages.

> *Yes, a bit, I have told my family, my friends as well, that "Stop nagging, it doesn't work." I think it is even worse then.*

Text Messages as a Complement

Respondents used the intervention as a complement to other smoking cessation strategies, such as medication and physical activity. As respondents were recruited by the smoking cessation specialist, some had continued support from the specialist, while others did not. Some respondents tried medication or nicotine patches with varied outcomes. Finally, other respondents described finding their own strategies to cope with abstinence, such as going for regular walks. Nevertheless, the text messaging, smoking cessation intervention was described as one piece of the jigsaw that was combined with all the above strategies.

> *Well...now when I was really motivated 'cause I had to stop, so I think that I could have done it without the text message, but it was a good complement to the sessions.*

Determinants of Use

Overview

The category *determinants of use* refer to aspects that influenced how and when patients used the intervention and included the timing of the intervention, motivation to change among patients, and perceptions of the mobile phone as a medium for support.

Timing

Timing referred to how well the intervention could respond to the needs of the patients. The timing of content contributed to how the intervention was perceived and used among patients. When the timing was optimal, a text message could offer support, consolidate or confirm feelings or states of the users, and empower respondents to keep going. However, when timing was suboptimal, text messages could prompt adverse experiences, such as irritation or craving. For example, one informant described that stopping smoking was not an issue or difficult to do, but that the intervention reminded him about cigarettes and prompted cravings. Another respondent described that she had already stopped smoking when she signed up for the intervention, which made the text messages about preparing to quit have limited relevance for her.

> *I will be totally honest; I didn't have any problems quitting at all. But then I got lots of text messages and then I got reminded of it and I started to feel and think, "Do I fancy a cigarette?" So it was the opposite for me. Do you know what I mean?*

Motivation to Change

The data showed that patients having elective surgery had a strong motivation and persistence for smoking cessation, which contributed to the openness toward using the intervention. There was an urgency to stop smoking among the respondents that stemmed from their health status or requirements from the surgery department. This strong motivation was illustrated by immediate action to stop smoking after having been given the information about surgery or diagnosis. One informant described that they finally had been given a reason to quit the cigarettes, and when compared to previous attempts to stop smoking, this time quitting was effortless.

> *And then I walked down to her at the unit straight away from meeting with the surgeon to speak with her. And then I said, "Let's throw away the cigarettes."*

However, the data also showed that motivation to stop smoking could be vulnerable and predominantly depend on surgery scheduling and outcome. For example, one informant described that they had to have multiple surgeries and that they timed their smoking cessation accordingly.

The respondents also took an active role in their care plan. Minimizing risks for complications after surgery were described as a team effort where the respondents acknowledged their responsibility to be smoke-free before surgery. Respondents described their appreciation about having access to surgery and felt that they needed to contribute to optimal conditions for surgery. For example, one respondent explained that she felt afraid of the risks of smoking before surgery and ashamed for being a smoker, and that these feelings contributed to her smoking cessation.

> *So, I think that when they are doing the operation and they have planned this and they want me to quit if I could, then I thought that "Of course I need to give to this as well."*
>
> *Yes, but absolutely. And especially that feeling that I...when I walked out of that room, like both scared and ashamed and...it became "Yes, let's do this."*

Perceptions of the Mobile Phone Medium

The data showed that users' perceptions of the mobile phone medium and the text messages used to reach them contributed to how they used the intervention and how they appraised the content. Respondents described that text messages as a medium to deliver the intervention was useful, reliable, and required little from them as users. For example, one respondent perceived text messages to be more personal than mobile apps, which were thought of as more generic. Another respondent expressed that text messages gave a more serious impression than other methods, such as social media.

> *So, it becomes like there is somebody that sits there and sends the message, or maybe they maybe it is automated, but you take it seriously 'cause it comes as a text message.*

Informants expressed that receiving cessation support and factual information on the phone, in the text message format, was described as valuable and that the content, albeit something they already knew, was thought about in a different way when it arrived via their personal phone.

> *I don't care what it says there [health adverts campaigns], but 'cause it was in a text message...yes, then it becomes more personal so you take it in a different way.*

However, although respondents described that they read most of the text messages, they expressed that they engaged very little with the interactive modules of the intervention. Respondents described that using the intervention had to be effortless.

> *And then it was like it was quite effortless for me, I don't need to do that much, but that there was somebody else all the time that reminded me or motivated me; do you know what I mean?*
>
> *I didn't want to sit with it [mobile phone]; I thought that it was so awkward with it. No, so...no, I didn't, I never...yes, that's what it's, it's awkward to sit. But I thought reading the text message, like, that I found ok.*

Discussion

Principal Findings

This study investigated the implementation of facilitated access where smoking cessation specialists invited patients to sign up for the intervention during a face-to-face visit. Also, the perceived usefulness of the text messaging, smoking cessation intervention among patients waiting to undergo surgery was explored. Smoking cessation specialists used strategies such as building trust with patients and adapting their approach to drive the implementation forward. Specialists perceived the intervention as a useful complement to routines and that its content and structure were compatible with existing practice. Findings from interviews showed that this patient group had a strong motivation to quit smoking and that the timing of the intervention influenced how text messages were perceived and used. Patients' use of the intervention was characterized by an openness to learn, embracing the advice given, and humanizing and sharing of text messages. The intervention was often used as a complement, in combination with other strategies such as physical activity.

Findings showed that smoking cessation specialists found the facilitated access elements relatively straightforward to implement into their routines. Specialists talked about how they recognized the content and advice given in the intervention to be compatible with their existing practice. Also, they perceived the intervention to be a valuable and useful complement to existing routines. Implementation theory and research propose that *how* a new practice is perceived among key stakeholders is important for how, and to what extent, it is implemented [32]. Accordingly, how users perceive the characteristics of an intervention in terms of its compatibility, trialability, complexity, relative advantage, and observability will contribute to how the intervention is used and implemented. Smoking cessation specialists expressed that the content of the intervention was compatible with other strategies that were used and that the intervention could fit into the session structure. However, other aspects, such as perceived complexity of the intervention and its implementation, relative advantage, and observability of the benefits of the intervention, could have made implementation more difficult. Findings do not indicate that facilitated access was difficult to carry out, per se; however, specialists expressed that they struggled to communicate the intervention to patients in an optimal way and that limited time made it challenging to always engage patients. Relative advantage and observability are difficult to pinpoint, as the intervention was thought of, and intended, as a complement to other support. Thus, implementing facilitated access of a text message intervention in routine smoking cessation practice seemed to have been facilitated by characteristics such as the perceived compatibility; however, implementation could have been further facilitated in terms of how the intervention was communicated to patients.

Furthermore, findings showed that specialists overcame barriers to implement by adapting how they presented and spoke about the intervention as well as which role they adopted, depending on the specific need of the situation or patient. Implementation theory proposes that innovations that can be adapted to fit local needs and resources are easier to implement. The notion of adapting evidence-based programs and interventions was defined decades ago as the degree to which an intervention is altered in the process of adoption and implementation [32]. Recent definitions highlight in what way adaptations occur, such as the content and delivery of an intervention [33,34]. Previous research has shown that adaptations occur due to the needs and situations of patients or practitioners [35], due to the restraints such as limited time or resources [35,36], to promote recruitment or retention [37], and to increase the fit between the intervention and the actual implementation setting [35]. A potential challenge with text message–based interventions, that are generic and automated, is that they can seldom be tailored to individual patients compared to, for example, motivational interviewing. However, our findings showed that the routines around the intervention and how practitioners talked about the intervention was adapted to the situation.

Moreover, adapting how the intervention was presented to patients meant that specialists took a relatively active role in the implementation process. The fact that they built trust with patients to contribute to introducing the intervention is another example of adopting an active role and taking responsibility for implementation. Interestingly, specialists used existing therapy strategies to enable and optimize the conditions for implementation (ie, building trust with patients). This suggests that facilitated access could be a useful setting for implementation, as it is facilitated by the patient-practitioner relationship. The way specialists approached implementation and their flexibility most probably facilitated implementation efforts.

Furthermore, the health situation of the patients could have promoted implementation and use of the intervention. The findings illustrated a patient group that was relatively engaged in both their behavior change and using the intervention. Patients were receptive of the intervention, which can be illustrated by their openness toward the content of the text messages and sharing of the intervention with significant others. However, how text messages were thought of and humanized indicate that the patients processed and integrated the content of the messages into their behavior change efforts. Self-determination theory proposes different types of motivations that are driven predominantly either by controlled conditions (eg, external motivation, where behavior change happens due to someone telling you to change) or autonomous conditions (eg, intrinsic motivation, where behavior change is rewarding in itself) [38]. Our findings suggested that the patient group expressed external motivation rather than intrinsic motivation. For instance, patients described that requirements to quit smoking before surgery and to optimize healing postsurgery had prompted and motivated them to quit smoking. Behavior change based on external motivation, rather than intrinsic motivation, is more difficult to sustain in the long term. Thus, smoking cessation before surgery could be an effective time to reach smokers; however, interventions should ensure that cessation persists is the long term.

However, the interviews also showed that usage was limited regarding interactive modules and depended on the components being effortless and on the timing, of both the intervention and the content, being optimal. The role of timing was highlighted by both specialists and patients. Patients stressed the importance of timing between their quit date and commencing the intervention. Practitioners suggested that identifying and recruiting patients would be more efficient in primary care rather than waiting for the surgery department to refer patients. Previous research has highlighted that preoperative interventions are indeed difficult to implement and that primary care is an ideal setting for implementing smoking cessation [39]. This has implications for future implementation efforts of similar interventions in terms of when to introduce similar interventions for smoking cessation support in health care organizations.

Methodological Considerations

A limitation of the study is the number of focus groups and interviews that were carried out. The study had access to two smoking cessation offices that, in total, employed three smoking cessation specialists. Although all the specialists were invited to focus groups, this is a limited number. In addition, only women took part in the focus groups, which could have been a potential limitation. All specialists that worked at the offices were invited to take part in focus groups and all were women. This may represent the real-world situation of smoking cessation and nursing practice, which includes professional contexts where women are overrepresented. Nevertheless, a limited number of men in the study could have been a limitation.

Credibility of the study could have increased by including additional hospitals. Unfortunately, this was not feasible within the time frame of the study. However, we believe that the findings are still valuable in illustrating how a text messaging, smoking cessation intervention can be implemented and used in this kind of setting. We also tried to endorse credibility by involving different researchers in the data collection (CL, PB, and KT), main data analysis (KT), and critical review of the main findings (all authors). The trustworthiness of the main findings in terms of credibility was reviewed by all authors. Trustworthiness of the study in terms of dependability was increased by using data from two perspectives (ie, specialists and patients), using interview guides for all data collection, and employing a structured and systematic data analysis process. Interestingly, findings from the two datasets verified each other, for example, regarding the importance of timing and the use of the intervention as a complement. Reflexivity was used during the data analysis to increase confirmability. This was done by critically reviewing the analysis process (eg, rationale for merging codes and creating categories). Confirmability could have been increased further, however, through also systematically using reflexivity during the data collection phase and creating an explicit audit trail depicting the whole research process. Finally, the findings could be transferable to smoking cessation counseling practice in hospital settings and among adult patient preoperative populations.

Implications

The findings showed that preoperative patients exhibit strong motivation for smoking cessation and an openness toward using their mobile phones in their efforts to quit smoking. These findings imply great potential for health care systems to incorporate digital tools in practice to support patients. However, more research is needed that investigates the impact of different implementation strategies. This study focused on facilitated access at a smoking cessation unit in a hospital setting; future studies could explore implementation in primary care settings. Indeed, the findings highlighted the importance of timing between patients' quit dates and the intervention start date. By exploring possibilities of reaching patients in primary care, the timing could potentially be improved.

Conclusions

Smoking cessation specialists adopted an active role in implementing the intervention by adapting their approach and fitting the intervention into existing routines. Patients showed strong motivation to quit smoking and an openness to incorporate the intervention into their behavior change journey; however, the timing of the intervention and messages were important to optimize support. A text messaging, smoking cessation intervention can be a valuable and feasible way to reach smoking patients having elective surgery.

Acknowledgments

The project was funded by the Kamprad Family Foundation for Entrepreneurship, Research and Charity (grant No. 20170008).

Authors' Contributions

PB had the original idea for the study and received funding. MB programmed the intervention and trial system for baseline and follow-up data collection. PB, CL, and KT carried out data collection. KT conducted the primary data analysis. KT and PB together wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

PB and MB own a private company that develops and distributes evidence-based digital lifestyle interventions to be used in health care settings. KT and CL declare no conflicts of interest.

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;380(9859):2224-2260. [doi: 10.1016/S0140-6736(12)61766-8]
2. Public health data. The Public Health Agency of Sweden. URL: http://fohm-app.folkhalsomyndigheten.se/Folkhalsodata/pxweb/sv/B_HLV/B_HLV__aLevvanor__aagLevvanortobak/hHLV_Tobaksvanor_alder.px/ [accessed 2020-08-17]
3. Agardh E, Moradi T, Allebeck P. Riskfaktorernas bidrag till sjukdomsbördan i sverige: Jämförelse mellan svenska och WHO-data [article in Swedish]. Lakartidningen. 2008. URL: https://lakartidningen.se/wp-content/uploads/OldWebArticlePdf/9/9006/LKT0811s816_821.pdf [accessed 2020-08-17]
4. Pearson RG, Clement RGE, Edwards KL, Scammell BE. Do smokers have greater risk of delayed and non-union after fracture, osteotomy and arthrodesis? A systematic review with meta-analysis. BMJ Open 2016 Nov 14;6(11):e010303 [FREE Full text] [doi: 10.1136/bmjopen-2015-010303] [Medline: 28186922]
5. Grønkjær M, Eliasen M, Skov-Ettrup LS, Tolstrup JS, Christiansen AH, Mikkelsen SS, et al. Preoperative smoking status and postoperative complications. Ann Surg 2014;259(1):52-71. [doi: 10.1097/sla.0b013e3182911913]
6. Mills E, Eyawo O, Lockhart I, Kelly S, Wu P, Ebbert JO. Smoking cessation reduces postoperative complications: A systematic review and meta-analysis. Am J Med 2011 Feb;124(2):144-154.e8. [doi: 10.1016/j.amjmed.2010.09.013] [Medline: 21295194]
7. Harden WD, Prochazka AV. Review: Preoperative behavioral interventions increase smoking cessation and reduce postoperative complications. Ann Intern Med 2014 Jul 15;161(2):JC2. [doi: 10.7326/0003-4819-161-2-201407150-02002] [Medline: 25023268]
8. Turan A, Mascha EJ, Roberman D, Turner PL, You J, Kurz A, et al. Smoking and perioperative outcomes. Anesthesiology 2011;114(4):837-846 [FREE Full text] [doi: 10.1097/aln.0b013e318210f560]
9. Quan H, Ouyang L, Zhou H, Ouyang Y, Xiao H. The effect of preoperative smoking cessation and smoking dose on postoperative complications following radical gastrectomy for gastric cancer: A retrospective study of 2469 patients. World J Surg Oncol 2019 Apr 02;17(1):61 [FREE Full text] [doi: 10.1186/s12957-019-1607-7] [Medline: 30940207]
10. Thomsen T, Villebro N, Møller A. Interventions for preoperative smoking cessation. Cochrane Database Syst Rev 2014;3 [FREE Full text] [doi: 10.1002/14651858.cd002294.pub4]
11. Gollust S, Schroeder S, Warner K. Helping smokers quit: Understanding the barriers to utilization of smoking cessation services. Milbank Q 2008 Dec;86(4):601-627 [FREE Full text] [doi: 10.1111/j.1468-0009.2008.00536.x] [Medline: 19120982]

12. Leitlein L, Smit E, de Vries H, Hoving C. Factors influencing Dutch practice nurses' intention to adopt a new smoking cessation intervention. J Adv Nurs 2012 Oct;68(10):2185-2194. [doi: 10.1111/j.1365-2648.2011.05903.x] [Medline: 22141413]
13. Smit ES, de Vries H, Hoving C. Determinants of practice nurses' intention to implement a new smoking cessation intervention: The importance of attitude and innovation characteristics. J Adv Nurs 2013 Dec;69(12):2665-2674. [doi: 10.1111/jan.12153] [Medline: 23600904]
14. Applegate B, Sheffer C, Crews K, Payne T, Smith P. A survey of tobacco-related knowledge, attitudes and behaviours of primary care providers in Mississippi. J Eval Clin Pract 2008 Aug;14(4):537-544. [doi: 10.1111/j.1365-2753.2007.00910.x] [Medline: 18462288]
15. Godin G, Bélanger-Gravel A, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviours: A systematic review of studies based on social cognitive theories. Implement Sci 2008 Jul 16;3:36 [FREE Full text] [doi: 10.1186/1748-5908-3-36] [Medline: 18631386]
16. Utvärderingsrapport: Tobaksfri Inför Operation [report in Swedish]. Linköping, Sweden: Landstinget i Östergötland; 2010. URL: https://vardgivarwebb.regionostergotland.se/pages/104495/Utv%C3%A4rderingsrapport%20Tobaksfri%20inf%C3%B6r%20operation.pdf [accessed 2020-08-17]
17. Tønnesen H, Faurschou P, Ralov H, Mølgaard-Nielsen D, Thomas G, Backer V. Risk reduction before surgery. The role of the primary care provider in preoperative smoking and alcohol cessation. BMC Health Serv Res 2010 May 12;10:121 [FREE Full text] [doi: 10.1186/1472-6963-10-121] [Medline: 20462417]
18. Scott-Sheldon LAJ, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: A systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
19. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: A meta-analysis. Prev Sci 2015 Feb;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
20. Haug S, Meyer C, Dymalski A, Lippke S, John U. Efficacy of a text messaging (SMS) based smoking cessation intervention for adolescents and young adults: Study protocol of a cluster randomised controlled trial. BMC Public Health 2012 Jan 19;12:51 [FREE Full text] [doi: 10.1186/1471-2458-12-51] [Medline: 22260736]
21. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin R, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
22. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of short message service text-based smoking cessation intervention among university students: A randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):321-328 [FREE Full text] [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
23. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: A meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
24. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: A systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
25. Mobile health (mHealth) for tobacco control. World Health Organization. URL: https://www.who.int/tobacco/mhealth/en/ [accessed 2020-08-17]
26. West R, Raw M, McNeill A, Stead L, Aveyard P, Bitton J, et al. Health-care interventions to promote and assist tobacco cessation: A review of efficacy, effectiveness and affordability for use in national guideline development. Addiction 2015 Sep;110(9):1388-1403 [FREE Full text] [doi: 10.1111/add.12998] [Medline: 26031929]
27. Müssener U, Linderoth C, Thomas K, Bendtsen M. mHealth smoking cessation intervention among high school students: 3-month primary outcome findings from a randomized controlled trial. PLoS One 2020;15(3):e0229411 [FREE Full text] [doi: 10.1371/journal.pone.0229411] [Medline: 32142514]
28. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. SMS-based smoking cessation intervention among university students: Study protocol for a randomised controlled trial (NEXit trial). Trials 2015 Apr 08;16:140 [FREE Full text] [doi: 10.1186/s13063-015-0640-2] [Medline: 25872503]
29. Müssener U, Bendtsen M, McCambridge J, Bendtsen P. User satisfaction with the structure and content of the NEXit intervention, a text messaging-based smoking cessation programme. BMC Public Health 2016 Nov 22;16(1):1179 [FREE Full text] [doi: 10.1186/s12889-016-3848-5] [Medline: 27876031]
30. Bendtsen M, Linderoth C, Bendtsen P. Mobile phone-based smoking-cessation intervention for patients undergoing elective surgery: Protocol for a randomized controlled trial. JMIR Res Protoc 2019 Mar 26;8(3):e12511 [FREE Full text] [doi: 10.2196/12511] [Medline: 30912757]
31. 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]
32. Rogers EM. Diffusion of preventive innovations. Addict Behav 2002 Nov;27(6):989-993. [doi: 10.1016/s0306-4603(02)00300-3]

33. Escoffery C, Lebow-Skelley E, Haardoerfer R, Boing E, Udelson H, Wood R, et al. A systematic review of adaptations of evidence-based public health interventions globally. Implement Sci 2018 Sep 26;13(1):125 [FREE Full text] [doi: 10.1186/s13012-018-0815-9] [Medline: 30257683]
34. Escoffery C, Lebow-Skelley E, Udelson H, Böing EA, Wood R, Fernandez ME, et al. A scoping study of frameworks for adapting public health evidence-based interventions. Transl Behav Med 2019 Jan 01;9(1):1-10 [FREE Full text] [doi: 10.1093/tbm/ibx067] [Medline: 29346635]
35. Cohen DJ, Crabtree BF, Etz RS, Balasubramanian BA, Donahue KE, Leviton LC, et al. Fidelity versus flexibility: Translating evidence-based research into practice. Am J Prev Med 2008 Nov;35(5 Suppl):S381-S389. [doi: 10.1016/j.amepre.2008.08.005] [Medline: 18929985]
36. Cooper BR, Shrestha G, Hyman L, Hill L. Adaptations in a community-based family intervention: Replication of two coding schemes. J Prim Prev 2016 Feb;37(1):33-52. [doi: 10.1007/s10935-015-0413-4] [Medline: 26661413]
37. Moore JE, Bumbarger BK, Cooper BR. Examining adaptations of evidence-based programs in natural contexts. J Prim Prev 2013 Apr 19;34(3):147-161. [doi: 10.1007/s10935-013-0303-6]
38. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
39. Vick CC, Graham LA, Henderson WG, Houston TK, Hawn MT. Translating preoperative smoking cessation interventions into routine clinical care of veterans: Provider beliefs. Transl Behav Med 2011 Dec;1(4):604-608 [FREE Full text] [doi: 10.1007/s13142-011-0096-1] [Medline: 24073083]

Abbreviations

NEXit: Nicotine Exit

Edited by C Dias; submitted 20.12.19; peer-reviewed by S Xuening, N Berninger, N Abed Elhadi Shahbari, M Nannarone, SR Ramos, W Ajiboye, L Zheng, G Williams; comments to author 24.02.20; revised version received 23.04.20; accepted 16.06.20; published 18.09.20.

Please cite as:
Thomas K, Bendtsen M, Linderoth C, Bendtsen P
Implementing Facilitated Access to a Text Messaging, Smoking Cessation Intervention Among Swedish Patients Having Elective Surgery: Qualitative Study of Patients' and Health Care Professionals' Perspectives
JMIR Mhealth Uhealth 2020;8(9):e17563
URL: https://mhealth.jmir.org/2020/9/e17563
doi:10.2196/17563
PMID:32945772

Original Paper

Design and Usability Evaluation of Mobile Voice-Added Food Reporting for Elderly People: Randomized Controlled Trial

Ying-Chieh Liu[1,2], PhD; Chien-Hung Chen[3], MSc; Yu-Sheng Lin[2], MD; Hsin-Yun Chen[4], MSc; Denisa Irianti[1], MSc; Ting-Ni Jen[5], BA; Jou-Yin Yeh[6], BSc; Sherry Yueh-Hsia Chiu[7,8], PhD

[1]Department of Industrial Design, College of Management, Chang Gung University, Taoyuan City, Taiwan

[2]Department of Internal Medicine, Health Promotion Center, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[3]Cybersecurity Technology Institute, Institute for Information Industry, Taipei, Taiwan

[4]Department of Nutrition Therapy, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[5]Department of Chemical and Materials Engineering, College of Engineering, Chang Gung University, Taoyuan, Taiwan

[6]Department of Health Care Management, College of Management, Chang Gung University, Taoyuan, Taiwan

[7]Department of Health Care Management and Healthy Aging Research Center, College of Management, Chang Gung University, Taoyuan, Taiwan

[8]Division of Hepatogastroenterology, Department of Internal Medicine, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan

Corresponding Author:
Sherry Yueh-Hsia Chiu, PhD
Department of Health Care Management and Healthy Aging Research Center
College of Management
Chang Gung University
No. 259, Wen-Hwa 1st Road, 333 Kwei-Shan
Taoyuan
Taiwan
Phone: 886 3 2118800 ext 5250
Email: sherrychiu@mail.cgu.edu.tw

Abstract

Background: Advances in voice technology have raised new possibilities for apps related to daily health maintenance. However, the usability of such technologies for older users remains unclear and requires further investigation.

Objective: We designed and evaluated two innovative mobile voice-added apps for food intake reporting, namely voice-only reporting (VOR) and voice-button reporting (VBR). Each app features a unique interactive procedure for reporting food intake. With VOR, users verbally report the main contents of each dish, while VBR provides both voice and existing touch screen inputs for food intake reporting. The relative usability of the two apps was assessed through the metrics of accuracy, efficiency, and user perception.

Methods: The two mobile apps were compared in a head-to-head parallel randomized trial evaluation. A group of 57 adults aged 60-90 years (12 male and 45 female participants) was recruited from a retirement community and randomized into two experimental groups, that is, VOR (n=30) and VBR (n=27) groups. Both groups were tested using the same set of 17 food items including dishes and beverages selected and allocated to present distinct breakfast, lunch, and dinner meals. All participants used a 7-inch tablet computer for the test. The resulting data were analyzed to evaluate reporting accuracy and time efficiency, and the system usability scale (SUS) was used to measure user perception.

Results: For eight error types identified in the experiment, the VBR group participants were significantly (P<.001) more error prone owing to the required use of button-tapping actions. The highest error rates in the VOR group were related to incomprehensible reporting speech (28/420, 6.7%), while the highest error rates in the VBR group were related to failure to make required button taps (39/378, 10.3%). The VOR group required significantly (P<.001) less time to complete food reporting. The overall subjective reactions of the two groups based on the SUS surpassed the benchmark and were not significantly different (P=.20).

Conclusions: Experimental results showed that VOR outperformed VBR, suggesting that voice-only food input reporting is preferable for elderly users. Voice-added apps offer a potential mechanism for the self-management of dietary intake by elderly users. Our study contributes an evidence-based evaluation of prototype design and selection under a user-centered design model. The results provide a useful reference for selecting optimal user interaction design.

Trial Registration: International Standard Randomized Controlled Trial Registry ISRCTN17335889; http://www.isrctn.com/ISRCTN17335889.

(JMIR Mhealth Uhealth 2020;8(9):e20317) doi:10.2196/20317

KEYWORDS

voice-added design; food report; elderly; usability evaluation; automatic speech recognition; mHealth; randomized trial

Introduction

Background

Older people are at increased risk of malnutrition [1-3], which can increase a range of health risks [4,5]. To prevent malnutrition in seniors, self-monitoring of food intake is a critical component of macronutrient intake assessment and calorie calculation [6]. Regular screening of dietary factors from food intake can help to identify individuals at risk of malnutrition [7,8]. Advances in digital technologies are creating new options for delivering health interventions to seniors [9,10], and powerful new software and hardware tools can support improved health maintenance through the collection, analysis, and interpretation of dietary intake data [11,12]. Digital dietary intake tracking has substantial potential for improving related health or nutrition outcomes [13]. Vogels [14] found that seniors actively adopted new technologies in their daily lives, and among these, mobile health apps offer many potential advantages including accessibility, scalability, and cost-effectiveness [15,16].

Challenges in Operation

Older adults typically experience reduced physical and physiological functioning that can increase the challenge involved in operating mobile apps, such as tapping buttons and scrolling down the screen [17], and such users often require app developers to provide enhanced interface usability to make operation faster and more accurate [16-21]. Recent technological advancements, such as voice and speech recognition [22,23], have added a new dimension to health app interface design for older adults [24]. While some studies have identified potential benefits from such technologies [25,26], there is little evidence of their effectiveness. Voice interactions are highly domain oriented. Previous research [24] stressed the need to provide evidence of effectiveness and found that, among commercial apps, diet and calorie tracking apps rarely use voice input for food intake reporting. Exploiting the potential benefits of such technologies depends on the degree to which they improve usability for the intended users. Effective interface development usually requires the active involvement of target end users throughout the development process [27].

Objectives

We previously provided a proof of concept for a combinatorial approach of dietary recording that accounts for a wide range of dish variations [28,29]. The concept was shown to have potential for use by seniors [28], but design feedback was needed to improve the speed and accuracy of dietary reporting for this target group. Therefore, this research integrated a voice-input design enhancement for mobile app–based food intake reporting. The design approach integrated both voice and typical button interactions in handheld devices to develop two distinct prototypes for testing and comparison. This study investigated the practical usage experience of both apps and assessed their relative effectiveness for use by older adults.

Methods

General Overview of the Approach

We developed our voice-added design based on a user-centered design model [30] through research, ideation, and implementation steps. A review of the relevant literature and commercial apps was conducted, and team members engaged in extensive brainstorming for broad design ideas. One idea raised in the ideation stage was to allow voice-based intake recording. Our previously reported effort involved one-time voice reporting of food ingredients, portion size, cooking method, and other food attributes of a single dish. The initial prototype was reviewed by two senior dietitians, and testing results showed that one of the major obstacles was misrecognition of speech inputs. To reduce system complexity, we applied a design heuristics approach for simplicity [31,32] to propose two alternatives. The first alternative, voice-only reporting (VOR), decomposes the food contents of a dish into two parts. The user would then use voice inputs to describe these major content items, with additional items added later using traditional touch screen or voice input. The second alternative, voice-button reporting (VBR), adds a voice input feature to the existing touch-screen input procedure, based on our previous combinatorial food reporting concept. Major food ingredients were reported by voice input, and the remaining ingredients were reported using traditional touch screen–based user interaction.

App Implementation

The two apps were implemented in the Android operating system for use on 7-inch tablet computers. The VOR app allows users to simultaneously verbally report food names and food attributes, whereas the VBR app allows users to verbally report food names and then select food attributes by clicking the optional buttons. The Google speech cloud service (Google, Inc) was used for continuous speech recognition in both apps. The developed interfaces included senior-friendly design elements, such as bigger buttons and text, a simple layout, and high-contrast colors. Based on recommended design guidelines for seniors [17], the two apps shared a common interface design, including placement of buttons, text, and icons. Clear and intuitive visual cues were used to facilitate user interaction.

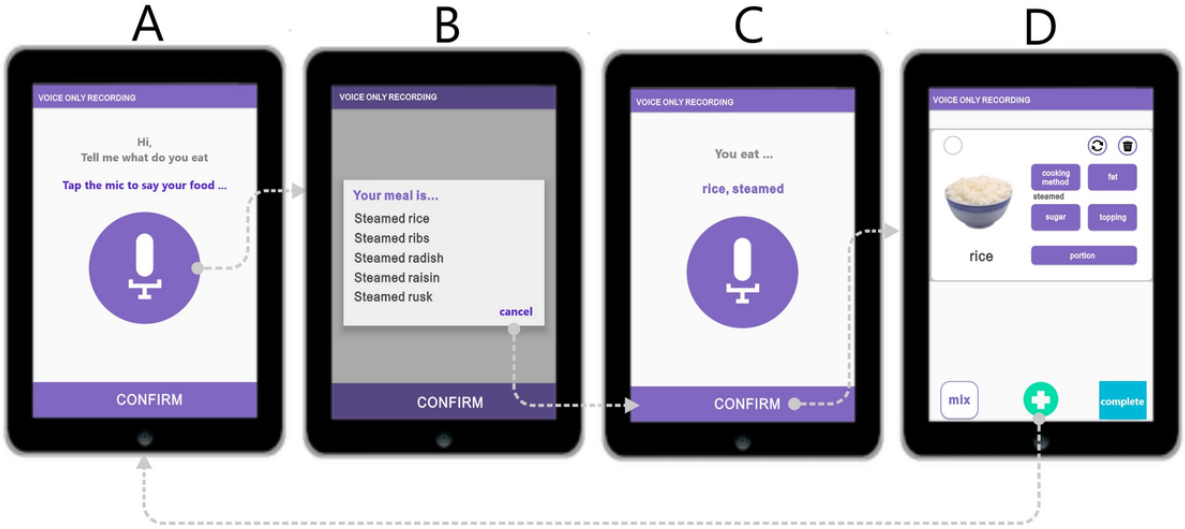
App Operation Overview

Voice-Only Reporting for a Dish With a Single Ingredient

Figure 1A shows the initial screen with the large "record" button used to activate speech recognition (Multimedia Appendix 1). The Chinese spoken language features many homophones, and recognized speech can be interpreted multiple ways. Figure 1B shows a menu of options resulting from the recognized speech, listed in descending order of confidence. The user selects the correct item from the list, and then, the system confirms the user's selection (Figure 1C). Thereafter, the confirmed selection is displayed (Figure 1D).

As shown in Figure 1B, the user selects the desired response from the listed responses or selects "cancel" to rerecord the input. A lexical filter prunes the list of potential responses to only eliminate less relevant responses, thus avoiding presenting the user with a long list of unlikely possibilities. A grammatical function called "Food Grammar" was developed to parse the selected results by selecting predefined keywords from the app's food name database based on four food attributes (method of cooking, sugar, fat, and topping). Words not kept by the filter are assumed to be food ingredients. For example, for "steamed rice," "steamed" is identified as the method of cooking, while "rice" is recognized as an ingredient. Following the parsing operation, the app displays the recognized food contents as "rice, steamed" (Figure 1C). When the user clicks "confirm" (Figure 1C), the screen presents a food editing page with the name of the food, along with corresponding images and food attributes (Figure 1D). Clicking the "add more (+)" button located at the bottom center (Figure 1D) allows the user to input additional ingredients or dishes by returning to the initial step (Figure 1A). As shown in Figure 1D, there is an attribute adjustment feature to account for variations in the method of cooking (eg, salad, boiled, stewed, stir fried, fried, and deep fried), sugar content (eg, 0%, 25%, 50%, 75%, and 100%), types of milk, topping (eg, tapioca bubbles, coconut milk, and ice cream), and portion size (eg, plate, bowl, cup, and other).

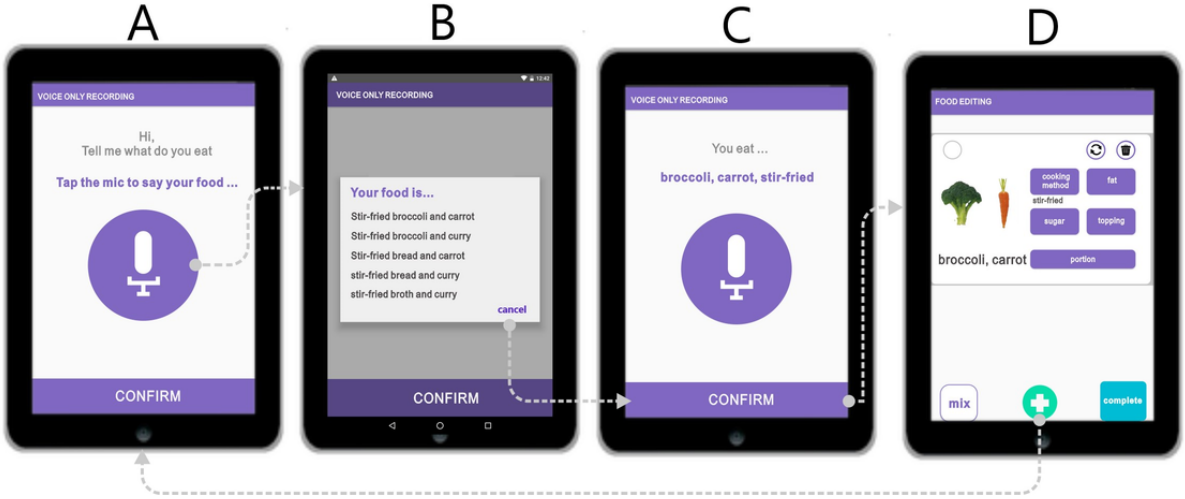
Figure 1. Voice-only reporting operation of a dish with a single ingredient, using steamed rice as an example.



Voice-Only Reporting for a Dish With Two or More Food Ingredients

For dishes with two ingredients, the user verbally inputs the first ingredient followed by its associated food attribute, and then repeats the process for the second ingredient (Multimedia Appendix 2). As shown in Figure 2A, the user inputs the dish as follows: "broccoli, stir-fried, carrot." In this case, the displayed list presents five possible alternatives (Figure 2B). The user selects the intended input from the list. After parsing, the selected result is shown (ie, broccoli, carrot, stir-fried; Figure 2C). The user then clicks "confirm" to move to the food editing page (Figure 2D). This input process is extended for dishes with two or more ingredients.

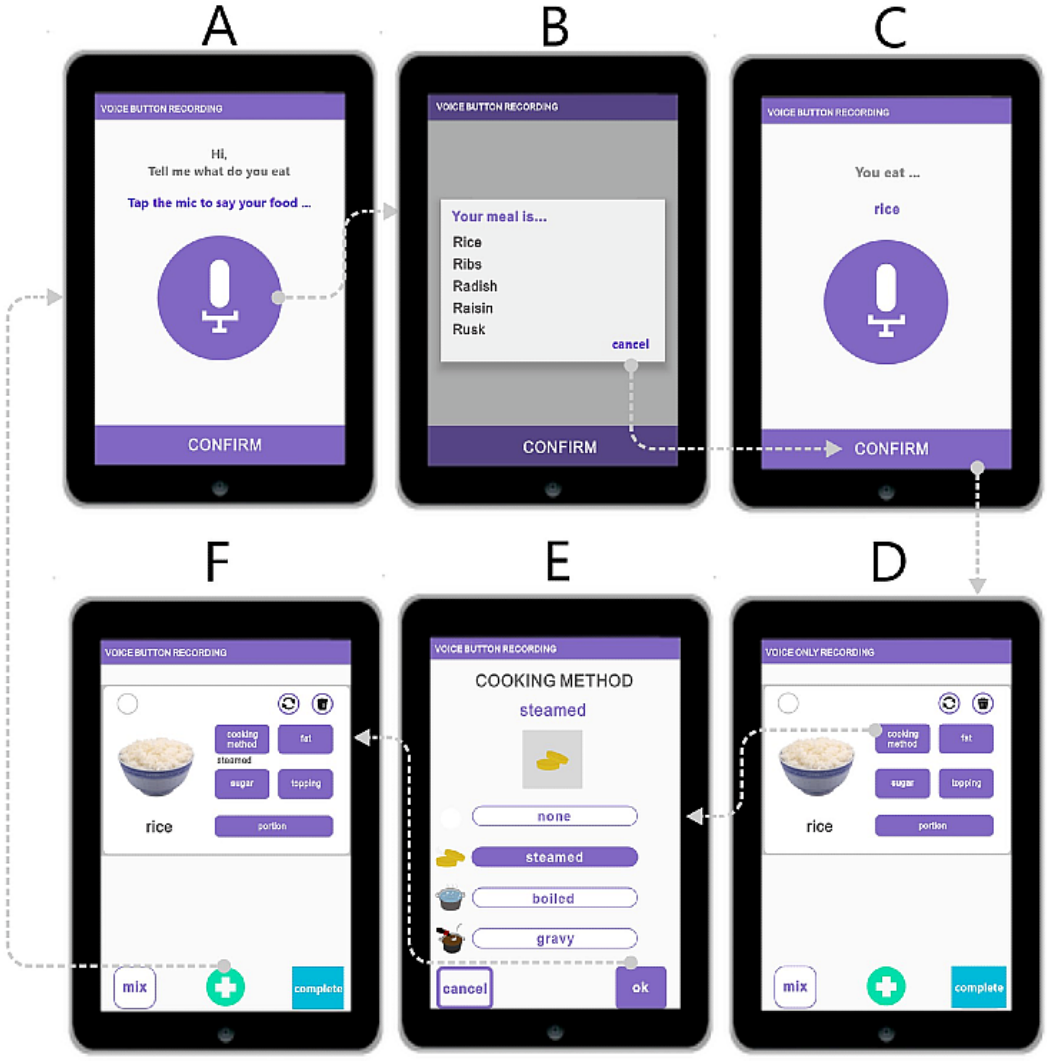
Figure 2. Voice-only reporting operation of a dish with two or three ingredients, using stir-fried broccoli with carrot as an example.



Voice-Button Reporting for a Dish With a Single Ingredient

Using VBR for a dish with a single ingredient, users first verbally input the name of the dish (Multimedia Appendix 3), for example, “rice” (Figure 3A). The food attributes are added in the following editing page by clicking one of the five buttons on the upper right (Figure 3D). The user thus adjusts the cooking method of “rice” by selecting “cooking method” (Figure 3D) and then selects the appropriate cooking method (eg, steamed; Figure 3E). The user clicks “OK” to then be presented with cooking method information (eg, steamed).

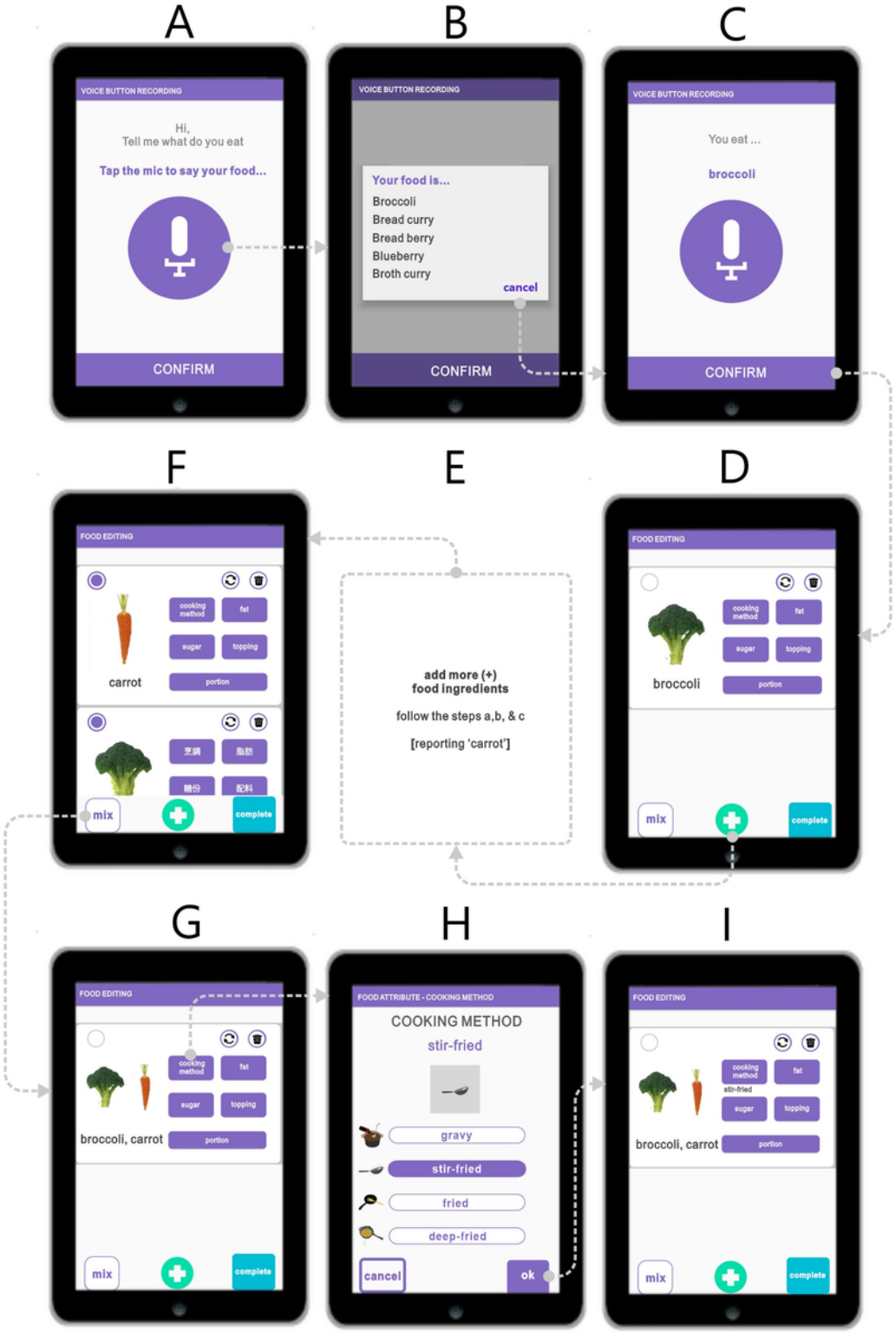
Figure 3. Voice-button reporting operation of a dish with a single ingredient, using steamed rice as an example.



Voice-Button Reporting for a Dish With Two or More Ingredients

For dishes with two or three ingredients, the user begins by verbally inputting the first food ingredient and follows the first four steps (Figure 4A-D; Multimedia Appendix 4). The user then clicks the "add more (+)" button and continues to report all subsequent food ingredients. When all ingredients have been reported, the user clicks the "mix" button to assemble the dish. As shown in Figure 4, the user reports "broccoli" (Figure 4D) and "carrot" (Figure 4E). The user selects the desired food items (eg, broccoli and carrot) and clicks the "mix" button (Figure 4F) to create mixed-food information (Figure 4G). The user then chooses the desired cooking method and clicks "confirm" (Figure 4H). The "add more (+)" button also allows the user to report other dishes while the "complete" button finishes the food reporting session (Figure 4I).

Figure 4. Voice-button reporting operation for a dish with two or more ingredients, using stir-fried broccoli with carrot as an example.



Study Design and Participant Recruitment

A parallel two-group randomized trial was designed to evaluate the relative effectiveness of the two apps in terms of reporting accuracy, task time, and user acceptance. The study protocol was reviewed by the Ethics Committee of Chang Gung Memorial Hospital and received Institutional Review Board approval (201900324B0). Recruitment was conducted through notices placed on designated bulletin boards in the Chang Gung Health and Culture Village retirement community located in northern Taiwan. Registration, schedule arrangement, and collection of background information were conducted through an online form. Biographic data were used to allocate participants into the VOR or VBR group. Self-reported baseline information included gender, age, BMI, experience in nutrition education, use of nutrition-related apps, cooking experience, and experience using mobile phones/tablets. Eligible participants were (1) aged from 60 to 90 years and (2) capable of reading and operating the app on their mobile phone. Participants currently under any form of dietary control, currently engaged in deliberate weight loss, or following a vegetarian diet were excluded. The assessment was conducted in a public area inside the community.

Dishes and beverages for the experiment were selected under the supervision of a senior nutritionist. The dishes were typical local Asian and Western-style foods. Three set meals involving 17 food items were used to represent breakfast, lunch, and dinner. Each set meal contained five food items (ie, a staple food, a main course, a dish with two ingredients, a dish with three ingredients, and a beverage). These set meals were presented on life-size colored food-photo boards (30 cm × 42 cm; photographed from above). Following previous research [28], to avoid disturbing variables, each dish was labeled (72 pt) above or below the food item (Multimedia Appendix 5).

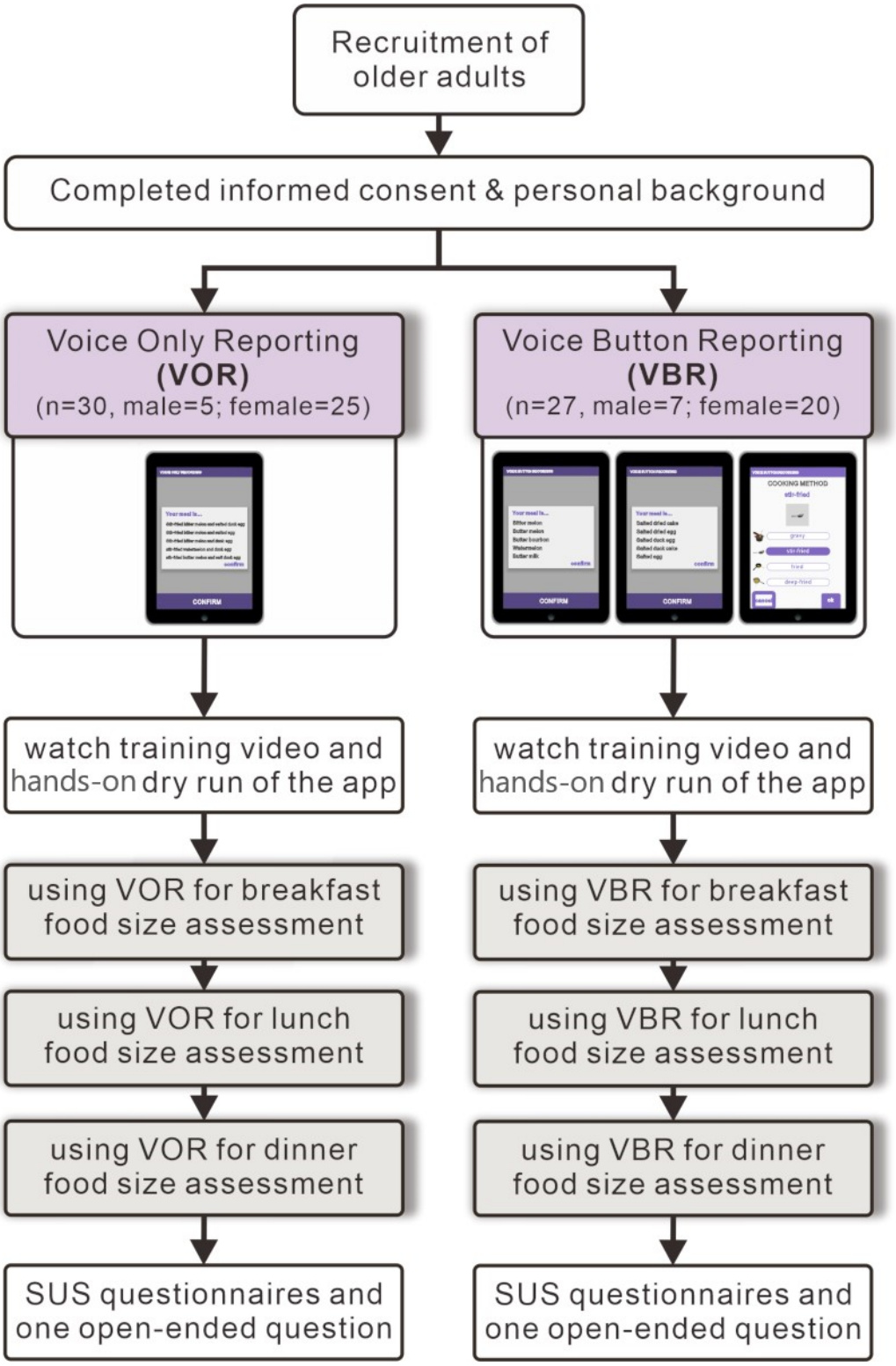
Sample Size Estimation

The sample size was based on our previous experience of customized dietary recording [28]. The mean difference for time required to complete the task using the two approaches was 4 seconds, with a SD of 5 seconds. Therefore, given a statistical power of 80% and a two-tailed α level of 5%, the minimum sample size required was 26 subjects each for VOR and VBR. Therefore, the minimum recruitment size was determined to be 52 subjects.

Randomization

A total of 57 senior participants were recruited and completed informed consent. SAS [33] was used to generate randomized lists of equal size with a 1:1 ratio for the two study arms, with 30 and 27 participants assigned to the VOR and VBR groups, respectively (Figure 5). The experiment was conducted with individual participants from each group in accordance with the randomization list.

Figure 5. App evaluation flow using a randomized design. SUS: system usability scale.



Evaluation Outcomes

The following three outcome types were assessed to evaluate the respective effectiveness of the two mobile apps for food reporting: accuracy, user operation time, and perception of efficacy.

Accuracy

An error was defined as the participant engaging in operating steps outside of those required to obtain the predefined answer. Possible error types of dish reporting were identified, and they have been described in the subsection “Error Types.” The rate of a specific error type was expressed as the error count divided by the total count. The error count was defined as the sum of participants with incorrect responses in the error type. The total count was calculated as the number of participants multiplied

by the number of all dishes. In a specific error type, each participant's reporting task might encounter more than one incorrect response, but was counted as one. The accuracy for error type was defined as the difference between the total count and error count divided by the total count.

Error Types

The error types were derived thematically for data analysis [34], which has been described in the subsection "Thematic Analysis." Data were derived from each participant-system interaction, with the app automatically logging the participant's selection of any functional buttons (eg, voice or mix button), and each interaction was tagged with a time stamp. The app also recorded the suggested results from the speech input along with the item subsequently selected and confirmed by the participant.

Task Duration

For VOR and VBR, the operating duration covered the time from when the participant began to input a food item until the participant tapped the "complete" button on the screen. For the VOR group, the task duration was calculated from the time the participant clicked the "voice" button to begin speaking to the time the participant clicked the "complete" button (Figure 1D and Figure 2D). For the VBR group, the task duration was calculated from the time the participant clicked the "voice" button to begin speaking to the time the participant completely mixed the multiple resulting ingredients and adjusted the cooking method (Figure 3F and Figure 4I). The mobile app automatically recorded the assessment duration of each participant.

Perception

The system usability scale (SUS) [35] was used to measure participant perception, with a questionnaire of 10 items. Each item used a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Following the study by Bangor et al [36], the mean SUS score with an adjective rating scale was used. Mean scores of 35.7, 50.9, 71.4, and 85.5 were rated as "poor," "ok," "good," and "excellent," respectively, in the adjective scales.

Assessment Procedures

The experiment was carried out by two research assistants. Informed consent was explained to and obtained from each participant. All participants utilized the same hardware (ie, a 7-inch Android tablet). All participant trials were conducted on a single day. Each participant session was scheduled by appointment and implemented individually. Each participant was first trained by watching an instructional video demonstrating how the food reporting app could be used. The researchers then spent several minutes teaching each participant how to navigate the interface, to ensure familiarity with app operation and features. The experiment was arranged on the basis of a set meal, with each set meal involving a staple food, a main course, a dish with two ingredients, a dish with three ingredients, and a beverage. Having understood the meal concept, each participant conducted a "dry run," which involved voice reporting of five food items (porridge, sausage, chicken egg, gluten with peanuts, and soy milk) on a photo board (Multimedia Appendix 5). A research assistant guided each participant to clearly pronounce the food name before conducting the food reporting. All participants were encouraged to become familiar with the app until they were confident with the app's operation flow. Participants were informed that their time to task completion was also a performance to be considered.

Respondents were asked to report three set meals (breakfast, lunch, and dinner). The first set meal, representing breakfast, featured boiled rice porridge, grilled pork sausage, stir-fried chicken egg, wheat gluten stewed with peanuts, and soy milk. The second set meal, representing lunch, featured steamed rice, deep-fried chicken, stir-fried broccoli with carrots, stir-fried tofu with green beans, stir-fried cabbage with bacon and black mushrooms, and green tea. The third set meal, representing dinner, featured fried noodles, pan-fried mackerel, stir-fried bitter melon with bell peppers and carrots, and tea with milk. The meal tests were performed in sequence, with a rest of 1 to 3 minutes between each test. The total test time for each participant took about 1 hour, beginning from when the participant first clicked the voice record button, according to the procedure shown in the "General Overview of the Approach" section. All participants completed the assessment.

Statistical Analysis

The chi-square test and *t* test were applied to examine the baseline characteristics of participants for categorical and continuous variables, respectively. According to our study endpoints, the accuracy between different groups was reported by the error proportion calculated as the number of errors by the total answer items. Second, the time duration for operating assessment was also employed for efficiency evaluation. As the time duration of reporting is a continuous variable, the *t* test was used to assess and compare the difference between the VOR and VBR groups. This comparison was also applied for dishes with different ingredients. SAS version 9.1.4 software (SAS Institute) was used to conduct all statistical analyses. All two-tailed statistical test results with a *P* value below .05 were considered to be statistically significant.

Thematic Analysis

Following the study by Bree and Galagher [37], all analyses were conducted using a Microsoft Excel worksheet (Microsoft Inc). The data were analyzed by two research assistants. The themes (ie, the error types) were first identified. The process of identifying a possible theme included the following steps. First, each assistant investigated the data independently, highlighting and labeling mismatched operating tasks for each dish. Similar labels were clustered into a single error type with a common tag, such as "missing food names" and "missing cooking methods" (Multimedia Appendix 6). Similar error types could be further grouped into a theme, such as "trouble after reporting" and "trouble in selecting one among the choices." When new error types or themes emerged, the overall network was revised accordingly. Discrepancies between the two assistants were discussed, and a consensus was reached under the supervision of the project leader.

Results

Participant Characteristics

A total of 68 participants were registered. Of these, 57 participants were scheduled and completed the experiment (Figure 5) and 11 failed to schedule an appointment. As shown in Table 1, 30 and 27 respondents were included in the VOR and VBR groups, respectively. The valid respondent pool had 21% (12/57) male and 79% (45/57) female participants, with an overall mean age of 73.92 (SD 1.48) years, where 12% (7/57) were aged 60 to 64 years, 40% (23/57) were aged 65 to 74 years, and the remaining 48% (27/57) were aged above 75 years. The mean BMI of all participants was 22.55 kg/m^2 (SD 2.25). Nearly half (28/57, 49%) of the participants had a bachelor's degree, followed by junior high school (12/57, 21%), senior high/vocational school (11/57, 19%), master's degree (5/57, 9%), and others (1/57, 2%). Looking at previous relevant experience, 95% (54/57) of respondents reported having experience in cooking, while 68% (39/57) reported previous experience using mobile phones or tablets, 32% (18/57) had taken nutrition-related courses, and 7% (4/57) had used nutrition-related apps. Among the two groups, the baseline information did not reveal relevant differences, confirming randomized allocation.

Table 1. Participant characteristics in the voice-only reporting and voice-button reporting groups.

Variables	Total (N=57), n (%) or mean (SD)	Voice-only reporting group (n=30), n (%)	Voice-button reporting group (n=27), n (%)	*P* value
Gender				.39
Male	12 (21%)	5 (17%)	7 (26%)	
Female	45 (79%)	25 (83%)	20 (74%)	
Age (years)[a]				.15
≤64	7 (12%)	6 (20%)	1 (4%)	
65-74	23 (40%)	10 (33%)	13 (48%)	
≥75	27 (48%)	14 (47%)	13 (48%)	
BMI (kg/m^2)[a]	22.55 (2.25)	22.63 (2.37)	22.45 (2.15)	.77
Education				>.99
Junior high school	12 (21%)	6 (20%)	6 (22%)	
Senior high/vocational school	11 (19%)	6 (20%)	5 (19%)	
Bachelor's degree	28 (49%)	14 (47%)	14 (52%)	
Master's degree	5 (9%)	3 (10%)	2 (7%)	
Others	1 (2%)	1 (3%)	0 (0%)	
Q1. Experience with nutrition-related courses				.40
Yes	18 (32%)	8 (27%)	10 (37%)	
No	39 (68%)	22 (73%)	17 (63%)	
Q2. Experience with health education				.58
Yes	17 (30%)	8 (27%)	9 (33%)	
No	40 (70%)	22 (73%)	18 (67%)	
Q3. Experience in cooking				.60
Yes	54 (95%)	29 (97%)	25 (93%)	
No	3 (5%)	1 (3%)	2 (7%)	
Q4. Experience using nutrition-related apps				>.99
Yes	4 (7%)	2 (7%)	2 (7%)	
No	53 (93%)	28 (93%)	25 (93%)	
Q5. Experience using mobile phones or tablets				.38
Yes	39 (68%)	19 (63%)	20 (74%)	
No	18 (32%)	11 (37%)	7 (26%)	

[a]Age and BMI data were analyzed with analysis of variance.

Error Types and Overall Accuracy

Table 2 presents overall accuracy results in terms of correct and incorrect counts for the 17 food items. Eight error types were identified and categorized from the analysis of each participant's food recording procedures (Multimedia Appendix 6). In the VOR group, error types were related to voice input (#1-3 and #5) and typical finger tap operations (#4 and #5-8). In the VBR group, error types were related to voice input (#1-3) and finger tap issues (#4-8). The error type "repeated pronunciations" (#5) was only relevant to VOR and "did not select the 'mix' button" (#7) only occurred in the VBR group. Among the eight error types, two error types ("missing cooking method(s)" [#4] and "did not select the 'mix' button" [#7]) showed significant differences (P<.001). The VOR group outperformed the VBR group in these two error types. In the VOR group, "no desirable choices" (#3) was the most commonly found error at a rate of 6.7% (28/420), followed by "incorrect selections in the list" (#6) at 3.8% (16/420) and "incorrect operation" (#8) at 2.6% (11/420). In the VBR group, "did not select the 'mix' button" (#7) was the most common error type at a rate of 10.3% (39/378), followed by "missing cooking method(s)" (#4) at 7.4% (28/378) and "no desirable choices" (#3) at 7.1% (27/378).

Table 2. Overall accuracy comparison of error types in the voice-only reporting and voice-button reporting groups.

Error type (correct/incorrect)[a]	Total (N=57), n (%)	Voice-only reporting group (n=30), n (%)	Voice-button reporting group (n=27), n (%)	*P* value
(#1) Missing the first food names/syllable(s)[b]				
Correct	796 (99.7%)	418 (99.5%)	378 (100.0%)	.50
Incorrect	2 (0.3%)	2 (0.5%)	0 (0.0%)	
(#2) Missing the last food names/syllable(s)[c]				
Correct	792 (99.2%)	416 (99.0%)	376 (99.5%)	.69
Incorrect	6 (0.8%)	4 (1.0%)	2 (0.5%)	
(#3) No desirable choices[d]				
Correct	743 (93.1%)	392 (93.3%)	351 (92.9%)	.79
Incorrect	55 (6.9%)	28 (6.7%)	27 (7.1%)	
(#4) Missing cooking method(s)[e]				
Correct	766 (96.0%)	416 (99.0%)	350 (92.6%)	<.001
Incorrect	32 (4.0%)	4 (1.0%)	28 (7.4%)	
(#5) Repeated pronunciations[f]				
Correct	794 (99.5%)	416 (99.0%)	378 (100.0%)	.13
Incorrect	4 (0.5%)	4 (1.0%)	0 (0.0%)	
(#6) Incorrect selections in the list[g]				
Correct	771 (96.6%)	404 (96.2%)	367 (97.1%)	.48
Incorrect	27 (3.4%)	16 (3.8%)	11 (2.9%)	
(#7) Did not select 'mix' button[h]				
Correct	759 (95.1%)	420 (100.0%)	339 (89.7%)	<.001
Incorrect	39 (4.9%)	0 (0.0%)	39 (10.3%)	
(#8) Incorrect operation[i]				
Correct	775 (97.1%)	409 (97.4%)	366 (96.8%)	.64
Incorrect	23 (2.9%)	11 (2.6%)	12 (3.2%)	

[a]Three items in beverage were not counted as no error types were found. Fourteen out of the 17 food items were included.

[b]#1 Missing first food name/syllable(s): After verbal reporting, the presented answer list did not include the first food name or the first syllable(s) of the food names.

[c]#2 Missing last food name/syllable(s): After verbal reporting, the presented answer list did not include the last food name or the last syllable(s) of the food names after voice reporting.

[d]#3 No desirable choices: After verbal reporting, the presented answer list did not present the desired food name or cooking method.

[e]#4 Missing cooking method(s): After verbal reporting, the presented answer list did not include the desired cooking method(s).

[f]#5 Repeated pronunciations: The presented answer list showed repeated pronunciations of food names and/or food attributes after voice reporting.

[g]#6 Incorrect selections in the list: Participant had trouble accurately tapping the desired choice (click interaction), leading to incorrect selection in the answer list.

[h]#7 Did not select the 'mix' button: Trouble before dish completion (click interaction). The user did not tap the "mix" button to complete dishes with two or three ingredients.

[i]#8 Incorrect operations: Incorrect operation procedure.

Accuracy and Trial and Error for Each Dish

The results are presented in terms of dish complexity (ie, number of ingredients) (Table 3). In addition, no errors were found for the three beverage items in either test group.

Table 3. Accuracy comparison of each food item in the voice-only reporting and voice-button reporting groups.

Food item and error type	Total (N=57), n (%)	Voice-only reporting group (n=30), n (%)	Voice-button reporting group (n=27), n (%)
Staple food			
Boiled rice porridge			
#1[a]	0 (0%)	0 (0%)	0 (0%)
#2[b]	0 (0%)	0 (0%)	0 (0%)
#3[c]	1 (2%)	0 (0%)	1 (4%)
#4[d]	3 (5%)	1 (3%)	2 (7%)
#5[e]	0 (0%)	0 (0%)	0 (0%)
#6[f]	0 (0%)	0 (0%)	0 (0%)
#7[g]	0 (0%)	0 (0%)	0 (0%)
#8[h]	0 (0%)	0 (0%)	0 (0%)
Steamed white rice			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	3 (5%)	2 (7%)	1 (4%)
#4	6 (11%)	0 (0%)	6 (22%)
#5	1 (2%)	1 (3%)	0 (0%)
#6	0 (0%)	0 (0%)	0 (0%)
#7	0 (0%)	0 (0%)	0 (0%)
#8	0 (0%)	0 (0%)	0 (0%)
Stir-fried noodle			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	0 (0%)	0 (0%)	0 (0%)
#4	2 (4%)	0 (0%)	2 (7%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	0 (0%)	0 (0%)	0 (0%)
#7	0 (0%)	0 (0%)	0 (0%)
#8	1 (2%)	0 (0%)	1 (4%)
Main course			
Grilled pork sausage			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	2 (4%)	1 (3%)	1 (4%)
#4	3 (5%)	1 (3%)	2 (7%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	2 (4%)	1 (3%)	1 (4%)
#7	0 (0%)	0 (0%)	0 (0%)
#8	2 (4%)	1 (3%)	1 (4%)
Stir-fried chicken egg			
#1	0 (0%)	0 (0%)	0 (0%)

Food item and error type	Total (N=57), n (%)	Voice-only reporting group (n=30), n (%)	Voice-button reporting group (n=27), n (%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	0 (0%)	0 (0%)	0 (0%)
#4	2 (4%)	0 (0%)	2 (7%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	3 (5%)	2 (7%)	1 (4%)
#7	0 (0%)	0 (0%)	0 (0%)
#8	1 (2%)	0 (0%)	1 (4%)
Deep-fried chicken leg			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	2 (4%)	0 (0%)	2 (7%)
#4	2 (4%)	0 (0%)	2 (7%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	3 (5%)	0 (0%)	3 (11%)
#7	0 (0%)	0 (0%)	0 (0%)
#8	1 (2%)	0 (0%)	1 (4%)
Pan-fried mackerel			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	10 (18%)	6 (20%)	4 (15%)
#4	2 (4%)	0 (0%)	2 (7%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	1 (2%)	1 (3%)	0 (0%)
#7	0 (0%)	0 (0%)	0 (0%)
#8	1 (2%)	1 (3%)	0 (0%)
Dishes with two ingredients			
Stewed wheat gluten with peanuts			
#1	1 (2%)	1 (3%)	0 (0%)
#2	1 (2%)	1 (3%)	0 (0%)
#3	19 (33%)	12 (40%)	7 (26%)
#4	3 (5%)	0 (0%)	3 (11%)
#5	2 (4%)	2 (7%)	0 (0%)
#6	12 (21%)	10 (33%)	2 (7%)
#7	14 (25%)	0 (0%)	14 (52%)
#8	6 (11%)	2 (7%)	4 (15%)
Stir-fried broccoli with carrot			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	1 (2%)	0 (0%)	1 (4%)
#4	1 (2%)	0 (0%)	1 (4%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	2 (4%)	0 (0%)	2 (7%)

Food item and error type	Total (N=57), n (%)	Voice-only reporting group (n=30), n (%)	Voice-button reporting group (n=27), n (%)
#7	3 (5%)	0 (0%)	3 (11%)
#8	0 (0%)	0 (0%)	0 (0%)
Stir-fried tofu with green bean			
#1	0 (0%)	0 (0%)	0 (0%)
#2	1 (2%)	1 (3%)	0 (0%)
#3	2 (4%)	1 (3%)	1 (4%)
#4	1 (2%)	0 (0%)	1 (4%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	0 (0%)	0 (0%)	0 (0%)
#7	7 (12%)	0 (0%)	7 (26%)
#8	1 (2%)	1 (3%)	0 (0%)
Stir-fried chicken egg with tomato			
#1	1 (2%)	1 (3%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	0 (0%)	0 (0%)	0 (0%)
#4	1 (2%)	0 (0%)	1 (4%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	1 (2%)	0 (0%)	1 (4%)
#7	5 (9%)	0 (0%)	5 (19%)
#8	2 (4%)	0 (0%)	2 (7%)
Stir-fried bitter melon with salted duck egg			
#1	0 (0%)	0 (0%)	0 (0%)
#2	1 (2%)	0 (0%)	1 (4%)
#3	7 (12%)	2 (7%)	5 (19%)
#4	1 (2%)	0 (0%)	1 (4%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	0 (0%)	0 (0%)	0 (0%)
#7	3 (5%)	0 (0%)	3 (11%)
#8	0 (0%)	0 (0%)	0 (0%)
Dishes with three ingredients			
Stir-fried cabbage with bacon and black fungus			
#1	0 (0%)	0 (0%)	0 (0%)
#2	1 (2%)	0 (0%)	1 (4%)
#3	5 (9%)	2 (7%)	3 (11%)
#4	4 (7%)	2 (7%)	2 (7%)
#5	0 (0%)	0 (0%)	0 (0%)
#6	2 (4%)	2 (7%)	0 (0%)
#7	5 (9%)	0 (0%)	5 (19%)
#8	5 (9%)	4 (13%)	1 (4%)
Stir-fried dry bean curd with bell pepper and carrot			
#1	0 (0%)	0 (0%)	0 (0%)
#2	2 (4%)	2 (7%)	0 (0%)

Food item and error type	Total (N=57), n (%)	Voice-only reporting group (n=30), n (%)	Voice-button reporting group (n=27), n (%)
#3	3 (5%)	2 (7%)	1 (4%)
#4	1 (2%)	0 (0%)	1 (4%)
#5	1 (2%)	1 (3%)	0 (0%)
#6	1 (2%)	0 (0%)	1 (4%)
#7	2 (4%)	0 (0%)	2 (7%)
#8	3 (5%)	2 (7%)	1 (4%)
Beverage			
Soymilk			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	0 (0%)	0 (0%)	0 (0%)
#4	0 (0%)	0 (0%)	0 (0%)
#5	0 (0%)	0 (0%)	0 (0%)
Green tea			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	0 (0%)	0 (0%)	0 (0%)
#4	0 (0%)	0 (0%)	0 (0%)
#5	0 (0%)	0 (0%)	0 (0%)
Milk tea			
#1	0 (0%)	0 (0%)	0 (0%)
#2	0 (0%)	0 (0%)	0 (0%)
#3	0 (0%)	0 (0%)	0 (0%)
#4	0 (0%)	0 (0%)	0 (0%)
#5	0 (0%)	0 (0%)	0 (0%)

[a]#1 Missing first food name/syllable(s): After verbal reporting, the presented answer list did not include the first food name or the first syllable(s) of the food names.

[b]#2 Missing last food name/syllable(s): After verbal reporting, the presented answer list did not include the last food name or the last syllable(s) of the food names after voice reporting.

[c]#3 No desirable choices: After verbal reporting, the presented answer list did not present the desired food name or cooking method.

[d]#4 Missing cooking method(s): After verbal reporting, the presented answer list did not include the desired cooking method(s).

[e]#5 Repeated pronunciations: The presented answer list showed repeated pronunciations of food names and/or food attributes after voice reporting.

[f]#6 Incorrect selections in the list: Participant had trouble accurately tapping the desired choice (click interaction), leading to incorrect selection in the answer list.

[g]#7 Did not select the ‘mix’ button: Trouble before dish completion (click interaction). The user did not tap the “mix” button to complete dishes with two or three ingredients.

[h]#8 Incorrect operations: Incorrect operation procedure.

Dishes With a Single Ingredient

These food items included three staple foods and four main courses. The VOR group featured fewer “missing cooking method(s)” errors (n=2) than the VBR group (n=18). The two groups showed similar results for error types #3 and #6. Both groups showed elevated error rates for error type #3 for pan-fried mackerel (n=6 in the VOR group; n=4 in the VBR group). In the VBR group, the incidence of error type #4 was higher for steamed white rice (n=6), but low for boiled rice porridge (n=2), stir-fried noodle (n=2), grilled pork sausage (n=2), stir-fried chicken egg (n=2), and deep-fried chicken leg (n=2). The incidences of other error types were relatively low.

Dishes With Two Ingredients

Five dishes included two ingredients. In the VOR group, error type #3 was more frequent for stewed wheat gluten with peanuts (n=12, 40%) and error type #6 was more frequent for stewed wheat gluten with peanuts (n=10, 33%). In the VBR group, error type #7 was more frequent for stewed wheat gluten with

peanuts (n=14, 52%), stir-fried tofu with green bean (n=7, 26%), and stir-fried chicken egg with tomato (n=5, 19%). In the VBR group, the frequency of error #3 was also relatively high for stir-fried bitter melon with salted duck egg (n=5, 19%). The incidences of other error types were relatively low in both groups.

Dishes With Three Ingredients

Three dishes were tested. In both the VOR and VBR groups, error type #3 occurred for two dishes, that is, stir-fried cabbage with bacon and black fungus (n=2 and n=3, respectively) and stir-fried dry bean curd with bell pepper and carrot (n=2 and n=1, respectively). Error type #7 had a higher incidence in the VBR group for stir-fried cabbage with bacon and black fungus (n=5), while error type #8 occurred frequently in the VOR group for stir-fried cabbage with bacon and black fungus (n=4). The incidences of other error types were relatively low in both groups.

Time Efficiency

Table 4 shows the time participants needed to complete the reporting task for the 17 food items. The results showed that the VOR group significantly outperformed the VBR group in terms of time efficiency (P<.001), with statistically significant advantages for all food items, aside from beverages. The per task operation time in the VOR group ranged from 9 to 42 seconds, as opposed to 8 to 70 seconds (mean 37.52 s) in the VBR group.

Table 4. Reporting time in the voice-only reporting and voice-button reporting groups.

Food item	Reporting time (s)			*P* value
	Total (N=57), mean (SD)	Voice-only reporting group (n=30), mean (SD)	Voice-button reporting group (n=27), mean (SD)	
Staple food				
Boiled rice porridge	20.44 (16.20)	10.50 (4.57)	31.49 (17.36)	<.001
Steamed rice	14.37 (8.19)	10.11 (4.98)	19.11 (8.53)	<.001
Stir-fried noodle	14.20 (10.75)	8.67 (3.45)	20.35 (12.69)	<.001
Main course				
Grilled pork sausage	26.39 (38.58)	12.20 (6.70)	42.15 (51.63)	.006
Deep-fried chicken egg	16.54 (10.47)	11.46 (8.82)	22.17 (9.32)	<.001
Fried chicken leg	16.80 (13.53)	8.99 (2.97)	25.48 (15.36)	<.001
Pan-fried mackerel	20.24 (12.69)	15.01 (11.23)	26.06 (11.81)	<.001
Dishes with two ingredients				
Stewed wheat gluten with peanuts	51.73 (32.09)	42.38 (28.05)	62.11 (33.58)	.02
Stir-fried broccoli with carrot	36.20 (36.30)	12.68 (4.17)	62.34 (38.34)	<.001
Stir-fried tofu with green bean	30.98 (27.86)	10.80 (4.90)	53.41 (25.56)	<.001
Stir-fried chicken egg with tomato	32.32 (28.73)	12.32 (6.32)	54.55 (27.54)	<.001
Stir-fried bitter melon with salted duck egg	33.90 (32.36)	12.39 (5.11)	57.80 (33.16)	<.001
Dishes with three ingredients				
Stir-fried cabbage with bacon and black fungus	44.39 (35.08)	21.23 (19.68)	70.13 (30.20)	<.001
Stir-fried dry bean curd with bell pepper and carrot	41.81 (33.76)	16.62 (8.42)	69.80 (28.82)	<.001
Beverage				
Soymilk	10.82 (7.20)	9.91 (6.31)	11.86 (8.11)	.31
Green tea	8.76 (2.66)	8.86 (3.07)	8.66 (2.16)	.78
Milk tea	9.64 (6.26)	10.81 (8.35)	8.35 (1.84)	.13

Time Efficiency for Dishes With One Ingredient

In the VOR group, the operation time ranged from 8 to 15 seconds per task, with pan-fried mackerel taking the longest time (mean 15.01, SD 11.23 s). In the VBR group, the operation time ranged from 19 to 41 seconds per task (mean 26.70 s), with grilled pork sausage taking the longest time (mean 42.15, SD 51.63 s). On average, the performance of the VOR group was roughly twice that of the VBR group.

Time Efficiency for Dishes With Two Ingredients

In the VOR group, four of the five dishes took 11 to 13 seconds, while stewed wheat gluten with peanuts took over 42 seconds. In the VBR group, the operation time ranged from 50 to 60

seconds, with stewed wheat gluten with peanuts taking over 60 seconds.

Time Efficiency for Dishes With Three Ingredients

The operation time in the VOR group ranged from 16 to 23 seconds, as opposed to 67 to 68 seconds in the VBR group.

Time Efficiency for Beverages

Both groups showed similar reporting operation time performance for beverages, with the VOR group taking 9 to 11 seconds per task, as opposed to 9 to 12 seconds in the VBR group.

System Usability Scale and Subjective Perception

Table 5 summarizes the SUS score and its two divisions in terms of usability and learnability. Overall scores showed no significant differences between the VOR and VBR groups (*P*=.20), but both exceeded the mean score of 71.4, indicating that the participants in both groups considered the app as being "good" to "excellent." In terms of learnability scores, the two groups showed a marginally significant difference (*P*=.06), suggesting that users found the VBR app slightly more difficult to learn to use.

Table 5. System usability scale and subjective perception in the voice-only reporting and voice-button reporting groups.

Score[a,b,c]	Voice-only reporting group (n=30), mean (SD)	Voice-button reporting group (n=27), mean (SD)	*P* value
Overall score	83.80 (9.49)	80.44 (10.25)	.20
Usability score	83.58 (9.57)	81.57 (9.69)	.43
Learnability score	84.67 (14.56)	75.93 (20.24)	.06

[a]Questionnaires were presented in Chinese.

[b]The mean score of the system usability scale with adjective ratings were as follows: 35.7 ("poor"), 50.9 ("ok"), 71.4 ("good"), and 85.5 ("excellent").

[c]The questionnaire's Cronbach α for voice-only reporting (α=.77) and voice-button reporting (α=.78) exceeded .70, indicating good internal consistency and reliability.

Discussion

Principal Findings

Two different voice-reporting designs were compared to investigate their respective effectiveness for food reporting among elderly users. VOR was designed to use verbal inputs for food names and attributes. VBR was operated through a sequential process of voice input and button tapping to report dietary intake. Experimental results showed the respective advantages and disadvantages of the two design concepts for authentic food reporting by older people. Our evidence-based findings provide insights into the relative usability of voice input in the food intake reporting context. The implications of these findings are discussed below, along with suggestions for further system improvement through the integration of voice input in the mobile health domain.

Accuracy Analysis: VOR Versus VBR

The eight error types identified in this research provide a useful reference for potential types of errors that will be encountered in voice-enabled user dietary intake interactions. The better performance of VOR for error types #4 and #7 indicates that VOR has the potential to provide greater accuracy in food reporting. Participants experienced error type #3 in both the VOR and VBR groups. This error is related to phoneme and syllable-based speech recognition issues, and food names or food attributes with similar phonemes tend to have lower recognition accuracy. For instance, in the VOR and VBR groups, error type #3 was most prevalent for "stewed wheat gluten with peanuts," and the Chinese term for "wheat gluten" (miàn cháng) was frequently misunderstood as "miàn chá." Participants also experienced a higher incidence of recognition errors for cooking methods, for example, lǔ (stew) was misrecognized as rǔ (milk), zhǔ (boil), and fǔ (rotten), contributing to the system's difficulty in accurately recognizing "stewed wheat gluten with peanuts." In addition, incorrect recognition results were found for food names such as "peanut," "bacon," and "salted duck egg," possibly because seniors have greater difficulty articulating nasal vowels [38]. One thing worth further investigating is the impact of the additional button tap required in VBR to identify food attributes. In contrast, VOR relied solely on verbal inputs and was designed to accommodate longer utterances, thus reducing the impact of error type #3.

The error type "did not select the 'mix' button" (#7) had the highest frequency among all errors in the VBR group and was specific to the item categories "dishes with two ingredients" (23.7%) and "dishes with three ingredients" (13.0%). This error may result from the app imposing cognitive overloading, as advanced age is associated with a decline in working memory [39]. VOR was designed to require fewer button taps, thus reducing opportunities for missing taps. To address this issue, future app designs should provide additional user training or design improvements [17] to prompt participants to remember to tap the required buttons.

The error "incorrect selections in the list" (#6) occurred with relatively high frequency in both groups (ranked second in the VOR group and third in the VBR group). This error is related to the user selecting the correct answer from a list of one to five possible choices, and could be explained by issues related to multimodal interaction in hand-eye coordination and speech input [40]. Our previous study [28] also addressed this issue for older adults, and system usability performance could be improved through improvements to the visual layout, increase in font size, or further improvements to user interaction design. Another multimodal interaction issue occurred in the error type "missing first food name/syllable(s)" (#1). However, this error occurred rarely. Among the 17 reporting task items, the three

beverage items did not incur any errors, suggesting that their relative simplicity made them relatively easy to report accurately in both apps.

Time Efficiency Analysis: VOR Versus VBR

For dishes with two ingredients, the need to tap buttons in the VBR group contributed to time efficiency up to five times worse than that in the VOR group (eg, 51.63 vs 10.67 s for "stir-friend tofu with green bean"). Aside from the beverage items, VOR consistently outperformed VBR in terms of time efficiency. The slower response time of VBR may be due to the need for additional button tapping to move between pages, and time spent on trial and error to obtain the correct food names or food attributes.

Participant Perception

The overall SUS score exceeded the adjective rating of "good," indicating that the participants considered the two apps to be useful. The high accuracy rate achieved by the two groups may conform with the high overall SUS scores. The significant time difference (P=.06) for task completion might reflect the marginally significant difference in the learnability component of the SUS model.

Use of Voice-Added Interfaces for Seniors

Some previous studies [17,26] suggested that senior citizens find button tapping-based interfaces on smartphones to be challenging. Following a previous report [28], the design of the apps developed for this study sought to simplify interface interaction as much as possible, to better suit the needs of senior users. The VOR interface only required one step for dish reporting, while the VBR interface required three steps (report food name, add food attribute, and mix the dish). In the VBR group, participants experienced errors in clicking food attributes (eg, error type #4, "missing cooking method(s)"). Additionally, some participants in the VBR group tended to forget or neglected the add food attributes and mix dish steps, thus reducing dietary intake reporting accuracy. However, neither of these steps was required in the VOR group. The situation is similar to reporting multiple ingredients or methods of cooking in that it requires multiple word inputs, but it did not reduce VOR performance. The results showed that VOR and VBR had a similar frequency of error type #3 for dishes with multiple ingredients. Moreover, VOR had significantly better time efficiency than VBR (P<.001). Previous research [25] has suggested that voice-enabled interfaces could potentially reduce barriers to use by elderly people having vision and motor disabilities. The voice-enabled interface provided in VOR optimizes this approach by minimizing button tapping requirements. Although voice-enabled interfaces may offer improved accessibility for older users, some issues still need to be investigated. This study found that participants using the voice input encountered recognition errors, and certain dish reporting tasks took a relatively long time to complete.

Limitations and Future Research

The experiments were conducted under laboratory conditions using a predetermined list of dishes and beverages. Participants were recruited from a retirement community; thus, further tests are required using different target populations (eg, seniors with specific chronic illnesses) whose results may differ from those of the groups tested here. The intended use case [41] in this research was to perform meal reporting with the use of the voice-added intake app, assuming users are familiar with the food ingredients and cooking methods of each dish. To better reflect realistic eating situations, future research should consider field user experience testing conducted in authentic settings. A wider range of authentic Asian and Western-style dishes and longer testing periods could also be included. Additional studies are needed to confirm the value of integrating voice inputs for food reporting. Further comparisons of the performance of voice-enabled and traditional interfaces under various eating contexts are needed. In addition, the idea of applying voice input to support existing dietary intake reporting apps could be explored to determine how and to what degree such integration improves usability. Further work also needs to include additional variables (eg, serving portion size, sugar and fat content, and toppings).

Conclusion

Experimental results showed that, while users assessed both VOR and VBR as having similar utility, VOR had better accuracy and time efficiency, making it a better candidate for food reporting by seniors. The design of VOR is superior to that of VBR in that it relies solely on voice input for food intake reporting and does not require additional button taps. Experimental results showed that speech recognition results for certain food items have reduced recognition accuracy, and both groups evidenced challenges in selecting the desired items from the postvoice input suggestion menu. The user experience assessment results for the two apps developed for this research provide a useful empirical reference for the development of high usability consumer apps for dietary monitoring among elderly people. Further studies are required, including investigations involving authentic dining environments with real-world meal options, along with full-scale randomized controlled trials to assess test efficacy.

Acknowledgments

The authors are grateful to all study participants at the Food Interaction Design Lab and the Health Informatics Lab of the Chang Gung University in the College of Management. This research was funded by the Research Fund of Chang Gung Memorial Hospital and Chang Gung University (BMRPD67 and BMRPB81) and the Ministry of Science and Technology, Taiwan (MOST-108-2221-E-182-008-MY3 and MOST-109-2314-B-182-038-MY3).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile app voice-only reporting of a dish with one ingredient.
[MP4 File (MP4 Video), 4005 KB - mhealth_v8i9e20317_app1.mp4]

Multimedia Appendix 2
Mobile app voice-only reporting of a dish with two or more ingredients.
[MP4 File (MP4 Video), 4235 KB - mhealth_v8i9e20317_app2.mp4]

Multimedia Appendix 3
Mobile app voice-button reporting of a dish with one ingredient.
[MP4 File (MP4 Video), 4557 KB - mhealth_v8i9e20317_app3.mp4]

Multimedia Appendix 4
Mobile app voice-button reporting of a dish with two or more ingredients.
[MP4 File (MP4 Video), 6315 KB - mhealth_v8i9e20317_app4.mp4]

Multimedia Appendix 5
Images of three set meals.
[PDF File (Adobe PDF File), 292 KB - mhealth_v8i9e20317_app5.pdf]

Multimedia Appendix 6
Error types of dietary intake using a voice reporting approach.
[DOCX File , 23 KB - mhealth_v8i9e20317_app6.docx]

Multimedia Appendix 7
CONSORT eHealth checklist.
[PDF File (Adobe PDF File), 31500 KB - mhealth_v8i9e20317_app7.pdf]

References

1. de Morais C, Oliveira B, Afonso C, Lumbers M, Raats M, de Almeida MD. Nutritional risk of European elderly. Eur J Clin Nutr 2013 Nov 25;67(11):1215-1219. [doi: 10.1038/ejcn.2013.175] [Medline: 24065060]
2. Volkert D. Malnutrition in older adults - urgent need for action: a plea for improving the nutritional situation of older adults. Gerontology 2013 Jul;59(4):328-333. [doi: 10.1159/000346142] [Medline: 23406648]
3. Clegg ME, Williams EA. Optimizing nutrition in older people. Maturitas 2018 Jun;112:34-38. [doi: 10.1016/j.maturitas.2018.04.001] [Medline: 29704915]
4. Keep fit for life: Meeting the nutritional needs of older persons. World Health Organization. 2002. URL: https://www.who.int/ageing/publications/keep_fit/en/ [accessed 2020-04-21]
5. Zhou X, Perez-Cueto F, Santos Q, Monteleone E, Giboreau A, Appleton K, et al. A systematic review of behavioural interventions promoting healthy eating among older people. Nutrients 2018 Jan 26;10(2):128 [FREE Full text] [doi: 10.3390/nu10020128] [Medline: 29373529]
6. McClung HL, Ptomey LT, Shook RP, Aggarwal A, Gorczyca AM, Sazonov ES, et al. Dietary intake and physical activity assessment: Current tools, techniques, and technologies for use in adult populations. Am J Prev Med 2018 Oct;55(4):e93-e104 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.011] [Medline: 30241622]
7. Katsas K, Mamalaki E, Kontogianni MD, Anastasiou CA, Kosmidis MH, Varlamis I, et al. Malnutrition in older adults: Correlations with social, diet-related, and neuropsychological factors. Nutrition 2020 Mar;71:110640. [doi: 10.1016/j.nut.2019.110640] [Medline: 31812000]
8. Robinson SM. Improving nutrition to support healthy ageing: What are the opportunities for intervention? Proc Nutr Soc 2018 Aug 27;77(3):257-264 [FREE Full text] [doi: 10.1017/S0029665117004037] [Medline: 29173227]
9. Bowles KH, Dykes P, Demiris G. The use of health information technology to improve care and outcomes for older adults. Res Gerontol Nurs 2015 Jan 01;8(1):5-10 [FREE Full text] [doi: 10.3928/19404921-20121222-01] [Medline: 25625621]
10. Samoggia A, Riedel B. Assessment of nutrition-focused mobile apps' influence on consumers' healthy food behaviour and nutrition knowledge. Food Research International 2020 Feb;128:108766. [doi: 10.1016/j.foodres.2019.108766]

11. Chen J, Gemming L, Hanning R, Allman-Farinelli M. Smartphone apps and the nutrition care process: Current perspectives and future considerations. Patient Educ Couns 2018 Apr;101(4):750-757. [doi: 10.1016/j.pec.2017.11.011] [Medline: 29169863]
12. Burrows TL, Rollo ME. Advancement in Dietary Assessment and Self-Monitoring Using Technology. Nutrients 2019 Jul 19;11(7):1648 [FREE Full text] [doi: 10.3390/nu11071648] [Medline: 31330932]
13. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
14. Vogels EA. Millennials stand out for their technology use, but older generations also embrace digital life. Pew Research Center. 2019 Sep. URL: https://www.pewresearch.org/fact-tank/2019/09/09/us-generations-technology-use/ [accessed 2020-04-21]
15. Ventola CL. Mobile devices and apps for health care professionals: Uses and benefits. Pharm Ther 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
16. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
17. Johnson J, Finn K. Designing User Interfaces for an Aging Population: Towards Universal Design. Cambridge: Morgan Kaufmann; Feb 16, 2017.
18. Levine DM, Lipsitz SR, Linder JA. Trends in seniors' use of digital health technology in the United States, 2011-2014. JAMA 2016 Aug 02;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
19. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: A usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
20. 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]
21. Wildenbos GA, Jaspers MW, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: Using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Apr;124:68-77 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
22. O'Brien K, Liggett A, Ramirez-Zohfeld V, Sunkara P, Lindquist LA. Voice-controlled intelligent personal assistants to support aging in place. J Am Geriatr Soc 2020 Jan 16;68(1):176-179. [doi: 10.1111/jgs.16217] [Medline: 31617581]
23. Hezarjaribi N, Mazrouee S, Ghasemzadeh H. Speech2Health: A Mobile Framework for Monitoring Dietary Composition From Spoken Data. IEEE J. Biomed. Health Inform 2018 Jan;22(1):252-264. [doi: 10.1109/jbhi.2017.2709333]
24. Chung AE, Griffin AC, Selezneva D, Gotz D. Health and fitness apps for hands-free voice-activated assistants: Content analysis. JMIR Mhealth Uhealth 2018 Sep 24;6(9):e174 [FREE Full text] [doi: 10.2196/mhealth.9705] [Medline: 30249581]
25. Cohen PR, Oviatt SL. The role of voice input for human-machine communication. Proc Natl Acad Sci 1995 Oct 24;92(22):9921-9927 [FREE Full text] [doi: 10.1073/pnas.92.22.9921] [Medline: 7479803]
26. Smith AL, Chaparro BS. Smartphone text input method performance, usability, and preference with younger and older adults. Hum Factors 2015 Sep;57(6):1015-1028. [doi: 10.1177/0018720815575644] [Medline: 25850116]
27. Chew S, Lai PS, Ng CJ. Usability and Utility of a Mobile App to Improve Medication Adherence Among Ambulatory Care Patients in Malaysia: Qualitative Study. JMIR Mhealth Uhealth 2020 Jan 31;8(1):e15146 [FREE Full text] [doi: 10.2196/15146] [Medline: 32003748]
28. Liu Y, Chen C, Tsou Y, Lin Y, Chen H, Yeh J, et al. Evaluating mobile health apps for customized dietary recording for young adults and seniors: Randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 15;7(2):e10931 [FREE Full text] [doi: 10.2196/10931] [Medline: 30767906]
29. Liu Y, Wu S, Lin S, Chen C, Lin Y, Chen H. Usability of food size aids in mobile dietary reporting apps for young adults: Randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 29;8(4):e14543 [FREE Full text] [doi: 10.2196/14543] [Medline: 32347805]
30. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
31. Cross N. Designerly Ways of Knowing: Design Discipline Versus Design Science. Design Issues 2001 Jul;17(3):49-55. [doi: 10.1162/074793601750357196]
32. Raab M, Gigerenzer G. The power of simplicity: A fast-and-frugal heuristics approach to performance science. Front Psychol 2015 Oct;6:1672 [FREE Full text] [doi: 10.3389/fpsyg.2015.01672] [Medline: 26579051]
33. Deng C, Graz J. Generating randomization schedules using SAS programming. 2002 Presented at: 27th Annual SAS Users Group International Conference; Apr 14-17, 2002; Orlando, FL URL: https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p267-27.pdf
34. Vaismoradi M, Jones J, Turunen H, Snelgrove S. Theme development in qualitative content analysis and thematic analysis. JNEP 2016 Jan 15;6(5):100-110 [FREE Full text] [doi: 10.5430/jnep.v6n5p100]

35. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Cambridge: Morgan Kaufmann; Jul 12, 2016.
36. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding and adjective rating scale. Journal of Usability Studies 2009 May;4(3):114-123 [FREE Full text]
37. Bree RT, Gallagher G. Using Microsoft excel to code and thematically analyze qualitative data: a simple, cost-effective approach. AISHE-J 2016 Jun;8(2):2811 [FREE Full text]
38. Bilodeau-Mercure M, Tremblay P. Age differences in sequential speech production: Articulatory and physiological factors. J Am Geriatr Soc 2016 Nov;64(11):e177-e182. [doi: 10.1111/jgs.14491] [Medline: 27783395]
39. Schroeder PJ. The Effects of Age on Processing and Storage in Working Memory Span Tasks and Reading Comprehension. Experimental Aging Research 2014 Apr 30;40(3):308-331. [doi: 10.1080/0361073x.2014.896666]
40. Stivers T, Sidnell J. Introduction: Multimodal interaction. Semiotica 2005 Dec 30;2005(156):1-20. [doi: 10.1515/semi.2005.2005.156.1]
41. Hartson R, Pyla PS. The UX Book: Agile UX Design for a Quality User Experience. Cambridge: Morgan Kaufmann; Nov 2, 2018.

Abbreviations

SUS: system usability scale
VBR: voice-button reporting
VOR: voice-only reporting

Edited by G Eysenbach; submitted 19.05.20; peer-reviewed by K Blondon, I Schiering; comments to author 01.07.20; revised version received 14.07.20; accepted 26.07.20; published 28.09.20.

Please cite as:
Liu YC, Chen CH, Lin YS, Chen HY, Irianti D, Jen TN, Yeh JY, Chiu SYH
Design and Usability Evaluation of Mobile Voice-Added Food Reporting for Elderly People: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(9):e20317
URL: http://mhealth.jmir.org/2020/9/e20317/
doi:10.2196/20317
PMID:32985999

Original Paper

Mobile Phone Apps for Food Allergies or Intolerances in App Stores: Systematic Search and Quality Assessment Using the Mobile App Rating Scale (MARS)

Floriana Mandracchia[1*], MS; Elisabet Llauradó[1*], PhD; Lucia Tarro[1,2], PhD; Rosa Maria Valls[1], PhD; Rosa Solà[1,2,3], MD, PhD

[1]Functional Nutrition, Oxidation, and Cardiovascular Diseases Group (NFOC-Salut), Healthy Environment Chair, Facultat de Medicina i Ciències de la Salut, Universitat Rovira i Virgili, Reus, Spain

[2]Unit of Nutrition and Health, EURECAT-Technology Centre of Catalonia, Reus, Spain

[3]Hospital Universitari Sant Joan de Reus, Reus, Spain

[*]these authors contributed equally

Corresponding Author:
Lucia Tarro, PhD
Functional Nutrition, Oxidation, and Cardiovascular Diseases Group (NFOC-Salut), Healthy Environment Chair
Facultat de Medicina i Ciències de la Salut
Universitat Rovira i Virgili
C/Sant Llorenç, 21
Reus
Spain
Phone: 34 977758920
Email: lucia.tarro@urv.cat

Abstract

Background: Food allergies and intolerances are increasing worldwide, and mobile phone apps could be a promising tool for self-management of these issues.

Objective: This study aimed to systemically search and assess food allergy or intolerance apps in app stores using the multidimensional Mobile App Rating Scale (MARS) to rate the objective and subjective quality and to identify critical points for future improvements.

Methods: This systematic search identified apps through the keywords “food allergy,” “food intolerance,” and “allergens” in English, Spanish, and Italian in the Apple App Store (iOS) and Google Play Store (Android). The inclusion criteria were a user star rating of ≥3 (of 5 stars) to limit the selection to the most highly rated apps; ≥1000 reviews as an indicator of reliability; and the most recent update performed up to 2017. Then, the apps were divided according to their purpose (searching for allergen-free “food products,” “restaurants,” or recipes in “meal planners”) and evaluated on a scale of 1 to 5 points using the MARS in terms of (1) app classification category with a descriptive aim; (2) app subjective and objective quality categories comprised of engagement, functionality, esthetics, and information sections (Medline was searched for eligible apps to check whether they had been tested in trials); and (3) an optional app-specific section. Furthermore, the output and input features were evaluated. Differences between MARS sections and between app purposes and correlations among MARS sections, star ratings, and numbers of reviews were evaluated.

Results: Of the 1376 apps identified, 14 were included: 12 related to food allergies and intolerances that detect 2-16 food allergens and 2 related only to gluten intolerance. The mean (SD) MARS scores (maximum 5 points) were 3.8 (SD 0.4) for objective quality, highlighting whether any app had been tested in trials; 3.5 (SD 0.6) for subjective quality; and 3.6 (SD 0.7) for the app-specific section. Therefore, a rating ≥3 points indicated overall acceptable quality. From the between-section comparison, engagement (mean 3.5, SD 0.6) obtained significantly lower scores than functionality (mean 4.1, SD 0.6), esthetics (mean 4, SD 0.5), and information (mean 3.8, SD 0.4). However, when the apps were compared by purpose, critical points were identified: meal planner apps showed significantly higher engagement (mean 4.1, SD 0.4) than food product (mean 3.0, SD 0.6; *P*=.05) and restaurant (mean 3.2, SD 0.3; *P*=.02) apps.

Conclusions: In this systematic search of food allergy or intolerance apps, acceptable MARS quality was identified, although the engagement section for food product and restaurant purpose apps should be improved and the included apps should be tested

in trials. The critical points identified in this systematic search can help improve the innovativeness and applicability of future food allergy and intolerance apps.

(JMIR Mhealth Uhealth 2020;8(9):e18339) doi:10.2196/18339

KEYWORDS

food allergy; food hypersensitivity; food intolerance; allergens; mobile applications; mobile health; mHealth; eHealth.

Introduction

Food allergies and intolerances are adverse reactions to the ingestion of, contact with, or inhalation of a specific food, derivative, or additive [1]. The prevalence of such adverse food allergy and intolerance reactions is increasing worldwide, especially in developed countries [2].

On the one hand, food allergies involve an immune-mediated reaction that occurs between a few minutes and 1 hour after exposure to the allergen, with symptoms ranging from moderate to severe [3]. The prevalence of food allergies is higher in children (<10%) than in adults (approximately 1%-2%) [3]. On the other hand, food intolerances are nonimmunological hypersensitivity responses due to a nontolerated dose of a food or a component of a food, with symptoms or signs occurring several hours after food consumption and lasting from hours until several days afterward [4]. Food intolerances are more common worldwide than food allergies, affecting up to 15%-20% of the general population [5].

Although new approaches to food allergies have recently been under clinical investigation [6], one strategy is to correctly identify food allergens to avoid the consumption of even small amounts of an allergen that causes a reaction [7]. To help consumers easily identify food allergens in food products, prepackaged or not, European legislation from 2014 (EU Food Information for Consumer Regulation No. 1169/2011) requires food businesses to clearly provide consumers, through labels or other verbal or written communications, with information about nutritional values and the presence of any of 14 specified food allergens (cereals containing gluten, crustaceans, eggs, fish, peanuts, soya, milk, nuts, celery, mustard, sesame, sulfur dioxide, lupin, and mollusks) [8]. Despite the European legislation, a 2019 study showed gaps in compliance with the regulation, finding that only 83 of the 295 evaluated restaurants (28.1%) labeled food allergens on the menus and that the restaurant staff had deficiencies in their food allergen knowledge and management [9]. In addition to relying on the information provided by food businesses and their employees, consumers must fundamentally self-monitor and self-manage their health [10].

In this context, there is increasing interest in mobile technology, such as apps, that focuses on helping consumers supervise what they are eating [11] by detecting allergens [12] not derived from cross-contamination and delivering specific health information [13,14] in relation to preparing daily meals, purchasing suitable food products, or searching for restaurants with allergen-free menus.

In recent years, mobile health (mHealth) technologies, including software, sensors, and mobile phones [15], have improved the management of health care services [16] and interventions such as the achievement of weight loss and smoking cessation as well as the management of several chronic and mental diseases [17]. Currently, the potential of apps for food-related conditions [18] such as food allergies and intolerances, whose incidence is growing worldwide [19], is also being studied. The convenience of apps in health management is favored by approximately 59% of the world population, corresponding to 4.57 billion people, mostly in northern Europe and the United States, who were active internet users in 2020 [20]. Apps enabling consumers to identify food allergens in foods and products, find allergen-free restaurants, and report and evaluate symptoms related to food allergies are already available, but most of them offer irrelevant and poor content [21].

Since plenty of apps currently exist, their reliability must be verified [22], as the traditional systems used to test app quality, such as users' star ratings (evaluating apps on a scale of 1 to 5 stars) and reviews, could allow fake or subjective reviews, giving wrong indications to users [23]. Furthermore, app descriptions in app stores are often incomplete or incorrect and are not a valid tool for assessing the quality of an app [24], especially when dealing with sensitive topics such as food allergies.

The necessity of regulating the quality and safety of mHealth technologies, defined by the World Health Organization as medical and public health practices supported by mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices [25], is particularly important for apps intended to be used for the diagnosis, cure, mitigation, treatment, and prevention of a disease or other conditions by aiding clinical decision-making [26]. These kinds of apps are classified and regulated as medical devices by the US Food and Drug Administration to ensure the safety of apps that are recommended by health professionals to their patients [26]. For instance, in 2015, the government of Catalonia (Spain) introduced a public platform for apps with quality accreditation from health professionals (mConnecta platform), thus establishing a safe and reliable environment for people to use these mHealth apps to self-monitor their health practices [27]. However, nonmedical apps intended to provide information and education to users, such as apps for food allergies and intolerances, also need to be regulated since incomplete information is often provided [28]. In this way, apps will provide better information to help users make health-related choices [29], mHealth will have more value, and fewer ineffective and unsafe apps will be available [30].

Owing to the necessity of ensuring better app quality for users, a Mobile App Rating Scale (MARS) was developed by a multidisciplinary team of experts as a simple, objective, and reliable tool for researchers, developers, and health professionals

to assess app quality and provide suggestions for future designs [31]. The MARS tool provides a multidimensional evaluation of app quality, whereas other existing tools mostly use one-dimensional measures. For example, the Intercontinental Medical Statistics Institute for Healthcare Informatics tool [32] assesses only app functionality, and the criteria of the Health Care Information and Management Systems Society tool [33] evaluate only app usability. The MARS tool has already been used for the quality assessment of different apps related to nutrition [34-36], sleep management [37], food provision [38], calorie counting [39], smoking cessation [40], physical activity [41], and weight management [42] but has not previously been used for food allergy or intolerance apps.

The aim of this paper was to systemically search app stores for apps about food allergies or intolerances, to assess the apps using the multidimensional MARS ratings of objective and subjective quality, and to identify the critical points for future improvements of these apps.

Methods

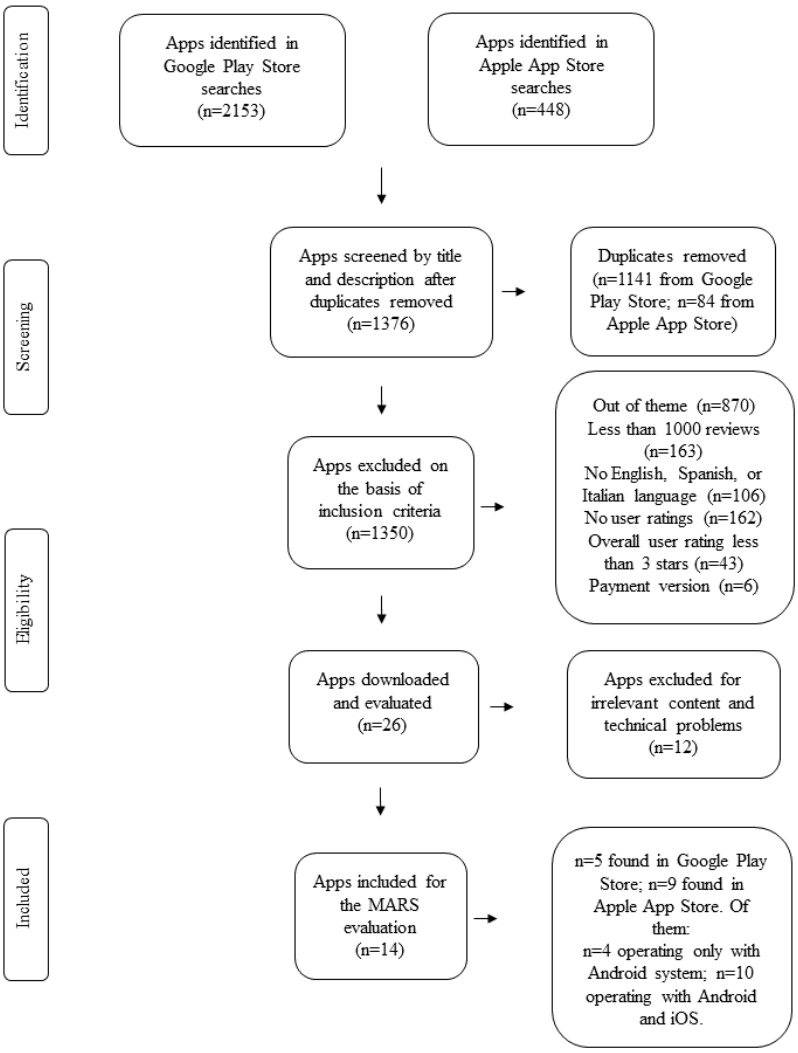
Search Strategy

The present study featured a systematic search and content analysis of apps about food allergies or intolerances available in the Apple App Store (iOS) and Google Play Store (Android). The apps were searched by the two authors between May 2019 and June 2019. The searches were conducted anonymously by logging out of the user accounts for the stores. Specific keywords such as “food allergy,” “food intolerance,” and “allergens” in English, Spanish, and Italian were used to search for the available apps in any of these 3 languages.

App Selection

The app selection process is described in Figure 1. Specific inclusion and exclusion criteria were applied to limit the search to the most relevant and reliable apps, in line with previous studies [35,38,41,42]. In particular, only apps that offered a free version were included in the search, as they are most commonly used by the general population. Apps in English, Spanish, and Italian were considered if they had (1) a minimum user star rating ≥3 (of 5 stars) to limit the search to the apps most highly rated by users, (2) ≥1000 reviews to identify the apps that were most commonly used and experienced, and (3) a last update up to 2017 to evaluate the most recently produced and revised apps. Finally, apps were included if their aim was to help allergic or intolerant users select suitable food products to buy or consume, personalize their daily nutrition on the basis of their needs and food restrictions, detect allergens in recipes and food product labels, search for specific restaurants or supermarkets according to their needs, and obtain information and advice about allergen self-management. Duplicates and apps that did not fulfill the aforementioned inclusion criteria or did not work were excluded from the study.

Figure 1. Flow diagram for the selection process of the apps included in the study.



Data Extraction

All the identified apps were registered in an initial list to count the total number of apps and the number of duplicates. The general characteristics of the included apps were extracted from the information in the app stores, while the main app features were verified by the authors by using the app. Furthermore, the features were categorized as input and output features on the basis of whether the app content was created by the users or automatically generated.

After data extraction, the authors divided the apps according to 3 purpose types (considering that the apps included presented different purposes): (1) searching for allergen-free food products, (2) searching for restaurants offering menus adapted to different food allergies and intolerances, and (3) functioning as meal planners for suitable daily meals according to users' food allergies or intolerances. This division of the included apps allowed us to compare the MARS quality ratings among apps with a similar purpose and to provide suggestions for future app designs in line with this purpose.

Moreover, web-based searches on the Medline database were conducted by app name (Eat This Much, Fitberry, Mealime, Recetas Vegetarianas y Veganas, SideChef, Tasty, Mercadona, Mi Intolerancia Alimentaria, Open Food Facts, ¿Qué Puedo Comer?, Club VIPS, Find Me Gluten Free, Foster's Hollywood, and Happy Cow) and by "apps for food allergies and/or intolerances" to determine whether they had already been evaluated in scientific trials.

MARS App Quality Assessment

App quality was assessed using the MARS rating scale, a reliable tool with a high internal consistency (α=0.90) and an interrater reliability interclass correlation coefficient of 0.79 [31]. The following steps were taken. First, before assessing the app quality, the authors followed specific web-based training organized by the MARS developers [43], such as an exercise to better understand how to classify the apps. Then, to experience and test the functionality of the included apps, the two authors independently used each of the 14 apps for 1 month. Finally, the quality assessment was conducted in agreement between the two authors, and disagreements were resolved through discussion with a third author.

The MARS rating scale consists of 2 categories. The first is the app classification category, with 6 items of descriptive and technical information for each app: (1) descriptive information (name, number, and type of ratings for all versions; developer; version; cost; platform; description; update), (2) focus, (3) theoretical background and strategies, (4) affiliations, (5) age group, and (6) technical aspects (login, password protection, web access, app community, social sharing, and reminder functions). The second category is the app quality category, which is divided into objective and subjective quality. Objective quality has 4 sections (engagement, functionality, esthetics, and information) with 19 items, while subjective quality is comprised of 4 items, for a total of 23 items.

In addition to these 2 categories, there is an optional app-specific section with 6 items to collect further information about the perceived impact of the app on the user (awareness, knowledge, attitudes, intention to change, help-seeking, behavior change).

The app classification category was not rated since its purpose was only descriptive. Instead, to evaluate the app quality category, each item was scored on a 5-point rating scale from 1 to 5 (1: inadequate; 2: poor; 3: acceptable; 4: good; 5: excellent). For each app, the total mean score was the sum of the score of each item divided by the number of total items. The mean score of the 4 objective quality sections (engagement + functionality + esthetics + information) was calculated separately from that of the subjective and app-specific sections to strengthen the impartiality of the measure.

For each objective quality section, the maximum score was 25 points for engagement, 20 points for functionality, 15 points for esthetics, and 35 points for information, for a total of 95 points for objective quality. Subjective quality could reach a maximum of 20 points, and the app-specific section could reach a maximum of 30 points.

In addition to the objective and subjective quality ratings, the app-specific section was evaluated on the 5-point rating scale.

Statistical Analysis

Continuous variables of the scores obtained for each section of the MARS quality assessment, with the exception of the app classification category, are presented as the mean and SD. Categorical variables for the included apps and their input and output features are presented as percentages. Multiple comparisons between the 3 purposes of the included apps (food products, restaurants, meal planners), MARS scores, and user star ratings were performed and adjusted using the generalized linear model of the Bonferroni test. Correlations among the MARS scores, user star ratings, and number of reviews were analyzed using Pearson correlation coefficients (for normally distributed variables) and Spearman correlation coefficients (for not normally distributed variables), which were interpreted as strong or moderate according to previously published cutoff points [44]. The analysis was performed with SPSS Statistics version 25. Statistical significance was considered at $P \leq .05$.

Results

App Selection

Figure 1 shows the flowchart of the app selection process. After the removal of duplicates found in both stores, 1376 apps about food allergies or intolerances were screened by title and description by the two authors, resulting in 1350 apps being excluded on the basis of the inclusion criteria. To further evaluate the eligibility of their content, 26 apps were downloaded, and 12 of these were excluded by common agreement because of irrelevant content (apps from the same developer with equivalent features and findings) and technical problems. As a result, 14 apps about food allergies or intolerances were finally included in the study for quality assessment using the MARS tool; 5 of the 14 (36%) were found in the Google Play Store, and 9 of the 14 (64%) were found in the Apple App Store. Moreover, 4 of the 14 apps (29%) operated only on the Android system, and 10 of the 14 apps (71%) operated on both the Android and iOS systems. None of the included apps had previously been evaluated in scientific trials.

Data Extraction: App General Characteristics

The general characteristics of the 14 included apps about food allergies or intolerances, shown in Multimedia Appendix 1, are described in the following sections.

App Purpose

First, the 14 included apps were divided according to their purpose.

Of the 14 apps, 6 (43%) were meal planners, helping users search for and plan meals adapted to allergies or intolerances. In particular, 4 apps (Tasty, Recetas Vegetarianas y Veganas, SideChef, and Fitberry) propose food recipes that can be filtered by the users' allergies or intolerances and on the basis of personal preferences, such as cooking difficulty and type of meal, diet, and cuisine. The other 2 apps (Mealime and Eat This Much) are meal planners that allow weekly meals to be organized on the basis of personal preferences, dietary goals, and food restrictions, such as food allergies and intolerances. In this way, users can create a personal profile indicating allergens to eliminate from their diet and organize their daily or weekly diet plan, choosing among the dishes proposed automatically by the apps and filtering them by the selected allergen.

Of the 14 apps, 4 (29%) function as food product search tools, helping users search for suitable food products according to their food allergies and/or intolerances. In particular, 3 apps (Open Food facts, ¿Qué Puedo Comer?, and Mercadona) help users search, through barcode scanning or database searches, for the most suitable food by showing the allergens declared on the food product label and indicating the nearest place to buy them, and 1 app (Mi Intolerancia Alimentaria) is a calculator of food compatibility. According to the presence of an allergen, the user's individual tolerance of the food or meal is calculated and shown using a 3-color code alert system (red, orange, and green) according to whether the compatibility of the food is low, medium, or high.

Restaurant searches represented the main purpose of 4 of the 14 apps (29%), helping users search for restaurants that offer menus adapted for allergic or intolerant consumers. In particular, 1 app (Happy Cow) searches for gluten-free, vegetarian, and vegan restaurants, hotels, supermarkets, and caterers; 1 app (Foster's Hollywood) belongs to a popular restaurant chain and offers the possibility of looking at the restaurant's allergen-free menu by checking the available meals in advance; 1 app (Club VIPS) searches for the nearest locations of different restaurant chains with allergen-free options; and 1 app (Find Me Gluten Free) searches for restaurants with gluten-free options.

Operating System

Of the 14 apps, 10 (72%) operate on both the Android and iOS systems, and 4 (28%) operate only on the Android system.

Number of Reviews

The number of reviews of the included apps varied from 1013 to 48,597 reviews.

Languages Available

Of the 14 apps, 4 (28%) are available only in Spanish, 5 (36%) are available only in English, and 5 (36%) are offered in 3-130 different languages.

Actions

The included apps enable users to benefit from different actions for the daily management of food allergies or intolerances.

Focus

Of the 14 apps, 12 (86%) are related to food allergies or intolerances, while 2 (14%) deal with gluten intolerance only.

Allergens Detected

The included apps differed in the number of allergens detected. Specifically, 10 of the 14 apps for food allergies identified milk and eggs; 9 identified crustaceans, peanuts, and nuts; 8 identified fish and soya; 6 identified sesame, mustard, and sulfur dioxide; 5 identified celery and lupin; and 1 identified wheat and grain. In addition, all 14 of the apps for food intolerances identified gluten, 4 identified lactose, and 2 identified fructose, sorbitol, histamine, and salicylic acid.

Thus, 5 of the 14 apps (36%) detected all 14 allergens that must be declared in the European Union (cereals containing gluten, crustaceans, eggs, fish, peanuts, soya, milk, nuts, celery, mustard, sesame, sulfur dioxide, lupin, and mollusks) above other food allergens present on the food product label, 2 of the 14 apps (14%) detected only gluten, and 7 of the 14 apps (50%) detected 2-10 food allergens.

Input and Output Features

The app features were distinguished as output features (Multimedia Appendix 2), where content is automatically generated by the app, and input features (Multimedia Appendix 3), where content is inserted and created by the user.

The lowest-rated app in the objective quality category, Mi Intolerancia Alimentaria (mean MARS score 3.2, SD 0.5), has fewer output features (4 of the 20 features) than the apps scoring >4 points, which offer 12-15 of the 20 output features and also had the highest scores in the engagement section. The same situation occurred for the input features, with apps scoring >4 points offering 8-9 of the 9 input features.

According to the app purposes, the most used features for the meal planner apps were allergen detection, search filters, sending of reminders and notifications, shopping list creation, suggestions and tips, rating and reviewing possibilities, personal profile, creation of a favorites list, and social sharing. For the food product apps, the most used features were allergen detection, listing of ingredients and additives, personal profile, and rating and reviewing possibilities. For the restaurant apps, the most used features were allergen detection, search filters, prompts and discounts, geolocation, rating and reviewing possibilities, personal profile, and social sharing.

MARS App Quality Assessment

The MARS app classification category is the part that collects descriptive and technical information about the included apps. Descriptive data include general information (app name, rating of all versions, developer, number of ratings of all versions, version, cost of basic and upgraded versions, platform, description and last update, focus, theoretical background and strategies, affiliations, age group) and technical aspects present in the app description in the app store. Of these data, only the relevant aspects were extracted (focus, theoretical strategies, affiliation, age group, and technical aspects); they are described in Multimedia Appendix 4.

According to the MARS evaluation, the quality of the 14 included apps assessed in terms of the 4 objectives (engagement, functionality, esthetics, and information) and the one subjective section are shown in Multimedia Appendix 4. Additionally, the results of the optional app-specific section are included.

The overall mean (SD) MARS objective quality score, which allows the evaluation of the general app quality (maximum of 5 points), was 3.8 points (SD 0.4 points); thus, the quality of the 14 included apps was considered acceptable. The score of the subjective quality section was 3.5 points (SD 0.6 points), and that of the app-specific section was 3.6 points (SD 0.7 points).

In particular, the mean scores of the 4 single objective quality sections, from the highest to the lowest score, were as follows: functionality section, 4.1 points (SD 0.6 points); esthetics section, 4 points (SD 0.5 points); information section, 3.8 points (SD 0.4 points); and engagement section, 3.5 points (SD 0.6 points).

When the scores of the 6 MARS sections (4 objective, 1 subjective, and 1 app-specific) were compared, the score of the esthetics section (mean 4, SD 0.5) was significantly higher than that of the engagement section (mean 3.5, SD 0.6; P=.007), and the score of the functionality section (mean 4.1, SD 0.6) was significantly higher than that of the subjective quality section (mean 3.5, SD 0.6; P<.001). Moreover, the score of the information section (mean 3.8, SD 0.4) was significantly higher than that of the subjective quality (mean 3.5, SD 0.6; P=.002) and app-specific (mean 3.6, SD 0.7; P=.001) sections. No further significance was found in the other between-section comparisons.

Among the 3 app purposes (food products, restaurants, and meal planners), comparisons between the MARS sections, as shown in Table 1, were evaluated. The score of the engagement section was significantly higher for meal planner apps (mean 4.1, SD 0.4) than for the food product (mean 3.0, SD 0.6; *P*=.05) and restaurant (mean 3.2, SD 0.3; *P*=.02) apps. Furthermore, it emerged that for meal planner apps, the scores of the engagement (mean 4.1, SD 0.4; *P*=.04) and functionality (mean 4.3, SD 0.7; *P*=.02) sections were significantly higher than those of the subjective quality section (mean 3.9, SD 0.5), and the score of the functionality section was significantly higher than that of the esthetics section (mean 4.3, SD 0.3; *P*=.04). No further significance was found in the other between-section comparisons among the 3 app purposes.

Table 1. Differences in the mean MARS scores between app purposes.

Mean MARS[a] scores	Meal planners	Food products	Restaurants	*P* value[b]	*P* value[c]	*P* value[d]
Engagement	4.10	3.00	3.20	.05	.02	1.0
Functionality	4.29	4.12	3.69	1.0	.43	.96
Esthetics	4.28	3.67	3.83	.17	.46	1.0
Information	3.97	3.79	3.62	1.0	.79	1.0
Subjective quality	3.87	3.31	3.19	.46	.26	1.0
App-specific	3.97	3.46	3.25	.75	.35	1.0

[a]MARS: Mobile App Rating Scale.

[b]Comparison between meal planners and food products.

[c]Comparison between meal planners and restaurants.

[d]Comparison between food products and restaurants.

Additional Analysis

The relationships between MARS score quality and user star rating and number of reviews were determined using correlations (described in Table 2) and showed that the star ratings were significantly and strongly positively correlated with the MARS engagement section (r=0.69; *P*=.007) and app-specific section (ρ=0.79; *P*=.001). A moderate correlation was also found between MARS subjective (r=0.63; *P*=.01) and total objective quality (r=0.60; *P*=.02). However, no significant correlations were found between MARS sections and number of reviews.

Table 2. Correlation coefficients between MARS scores, user star ratings, and number of reviews.

Mobile App Rating Scale (MARS)	Number of reviews	Star ratings	*P* value[a]	*P* value[b]
Functionality[c]	0.13	0.33	.65	.25
Esthetics[c]	0.11	0.43	.71	.12
App-specific[c]	0.05	0.79	.87	.001
Number of reviews[c]	1.00	0.30	N/A[d]	.30
Engagement[e]	0.20	0.69	.50	.007
Information[e]	–0.14	0.42	.62	.14
Total objective quality[e]	0.03	0.60	.92	.02
Subjective quality[e]	–0.03	0.63	.93	.01
Star ratings[e]	0.30	1.00	.29	N/A

[a]Correlation between MARS scores and number of reviews.

[b]Correlation between MARS scores and star ratings.

[c]Spearman (ρ).

[d]N/A: not applicable.

[e]Pearson (*r*).

In addition, to verify whether the star ratings assessed by users were similar to the MARS scores obtained in our study, the comparisons were analyzed. The star ratings were significantly higher (mean 4.2, SD 0.4) than the MARS subjective quality score (mean 3.5, SD 0.6; *P*=.04).

Discussion

The present systematic search and quality assessment study provides information about the objective (engagement, functionality, esthetics, and information) and subjective quality

of the available apps for food allergies or intolerances in app stores. The quality assessment using the MARS tool indicated that the overall app quality of the 14 included apps was acceptable, according to MARS mean ratings of ≥3 from a maximum of 5 points.

By comparing the 6 MARS sections (4 objective quality, 1 subjective quality, and 1 app-specific), the most significant results were related to the apps' functionality, esthetics, and information, as they appeared visually pleasant, sufficiently descriptive, well arranged, and easy to use, whereas the engagement section of most of these apps needs to be improved. As observed in other studies, apps with simple functionality can motivate people who have no familiarity with technology to adopt mobile apps [45]. Moreover, esthetics, such as visual attractiveness, is another key element for increasing users' motivation to use the app [46].

Regarding the information section, the included 14 apps clearly presented their content through the support of images, graphics, and videos. Nevertheless, none of the apps has been tested in scientific trials, which is an important aspect of this section of the MARS tool. In addition, it is important to evaluate the apps' efficacy in helping consumers self-manage food allergies or intolerances, since previous studies have demonstrated that commercial apps do not always provide the expected results when they are evaluated in trials [47-49]. For meal planner apps, future trials could evaluate the improvement in user knowledge and awareness of food allergens, which are considered important targets for the management of food allergies [50].

Moreover, the efficiency of food product apps should be tested in clinical trials to increase users' confidence when food shopping and reading product labels. For the allergic and intolerant population, it is fundamental for the food labeling system to be available and comprehensive [51], and this kind of app could help consumers more quickly detect allergens in food products. Finally, for restaurant apps, customer satisfaction when eating away from home could be evaluated as a measure of food businesses' compliance with the European regulation and with the allergen-free menus published on the app. Positive experiences when eating away from home are correlated with the availability of food allergen information provided by the restaurants [52].

Moreover, none of the 14 included apps claims any validation of the content by health professionals or allows remote support. Actually, a critical assessment published in 2015 found that most apps about food allergies lack important health information and are not developed with the support of health professionals [21]. It is important for such apps to be evaluated by health professionals to provide better information to help users make health-related choices [53-55]. Furthermore, apps providing professionally oriented support and communication are more engaging and favored by users, especially adolescents [54].

The results obtained in the present study indicate that app engagement is the section with the lowest score with respect to functionality, esthetics, and information, in line with other MARS assessments of apps for food provision [38], checking for drug interactions [56], and drunk driving prevention [57], and the lack of interactive features influences the engagement quality of these apps. However, in a comparison of the 3 purposes of the included apps, the engagement section of the meal planner apps received higher quality scores than that of the food product and restaurant apps. In fact, food product and restaurant apps do not use interactive features that motivate users to use them repeatedly [58], but for these apps, which are designed for short and specific use such as finding restaurants or products, user engagement and daily use are not really as essential as in meal planner apps. However, including features such as tips and suggestions to support consumers' decisions or sending notifications [59] to notify users of new products or restaurants could improve the user app experience, growth of the app community, and app competitiveness. To increase user enjoyment and participation, meal planner, food product, and restaurant apps should perhaps include features such as rewards, goal-setting options, challenges, and leader boards, which have been recognized as effective tools in past studies [60-62], especially in adolescent populations, where game competition can motivate users to participate [63]. Finally, features such as feedback and self-monitoring, which have been demonstrated in previous studies to be effective in increasing users' motivation [32,58] and health behavior [64,65], should be available in apps focused on self-managing food allergies or intolerances; however, only 2 of the 14 apps included in the present systematic search offer these features.

The subjective quality and app-specific sections need to be improved in relation to the 3 purposes of the included apps (meal planners, food products, and restaurants). These sections refer to general users' impressions of the app, which, if positive, would lead them to recommend and use it. In this context, the lack of enough engagement could influence users' perceptions. Thus, it is important to increase users' subjective quality perception and impact of the apps (app-specific) by reinforcing, for example, the engagement profile, as discussed earlier, which mainly influences users' view of the app.

Based on the number of input and output features offered, among the meal planner apps, Eat This Much, Mealime, and SideChef were found to be the most practical for users, obtaining higher scores in the MARS assessment than other apps with the same purpose. Previous studies have shown that food allergies and intolerances impact people's quality of life and emotional status, increasing anxiety and depression [66,67]. The avoidance of food allergens requires constant attention because their presence in food is not always evident or is unknown [68]. This problem becomes even more complicated when consumers have to adapt food recipes or make appropriate ingredient substitutions according to their allergy or intolerance [69] without accurate recommendations or support. In this sense, these 3 apps could better help users while providing suggestions for self-managing food allergies or intolerances in terms of cooking and daily menus. Among the food product apps, the ¿Qué Puedo Comer? app was the most practical for users, offering more features and gaining higher scores in the MARS assessment. This app helps consumers understand food product labels, detect food allergens, and search for food products according to allergies, intolerances, or dietary requirements. Since food product ingredients change regularly and consumers may need to read packaging labels several times [69], these apps can provide instant information

and support [70]. Among the restaurant apps, the Find Me Gluten Free and Happy Cow apps were the most practical for users compared to others with the same purpose. The provision of food allergen information on restaurant menus is very important for consumers, and these kinds of apps encourage the dissemination of such information by making it easier to search for restaurants with allergen-free menus [52].

The correlations of star ratings with the app-specific, engagement, and subjective sections suggest that when evaluating an app, users refer more to the subjective impression of the app given by the engaging features offered than to the quality and quantity of the information provided [71], as shown by the results obtained in the present study. As observed in previous studies, there is an evident difference between the quality evaluation obtained by a researcher using a more objective tool such as the MARS and a real-world user who tends to evaluate app quality through star ratings in a much more subjective way [72]. Nevertheless, app store user star ratings cannot be totally trusted since they are sometimes derived from piloted reviews or paid bots deployed by the developer [73].

Thus, according to the results obtained, we consider that MARS quality assessment is a valid tool for providing more accurate app quality information and suggestions for future apps.

Suggestions for Future App Development

Based on the present app assessment, several suggestions emerged for the future design of high-quality apps focused on improving the wellbeing of subjects with food allergies or intolerances:

1. Further features should be included, especially in meal planner apps, to improve the user app experience and increase participation.
2. Content should be validated by health professionals and scientists to provide users with more reliable information about food allergies or intolerances [36].
3. Remote support by health professionals would help users manage their food allergies or intolerances [54].
4. Testing in scientific trials would demonstrate the apps' reliability and effectiveness [74] in detecting food allergens and improving user knowledge.
5. Regulation of nonmedical apps should be considered in the future since it would avoid the development of unrealistic and ineffective apps, provide more correct information to users [29], and provide more value to mHealth technology [30].
6. App quality should be evaluated through innovative methods, including multiple dimension perspectives, as in the MARS tool. The MARS tool, compared to other scales [32,33], represents a multidimensional evaluation of app subjective quality as well as engagement, functionality, esthetics, and information as indicators of objective quality. However, although the MARS tool has been widely tested, it should be validated in the near future [75] to increase its value, and, depending on the area of interest of the app (eg, health care, nutrition, sports, psychology), the items in each section should be more specific and theme-based. Apps for food allergies or intolerances, for example, should include items asking whether food allergen information is effectively and appropriately provided to users.

Limitations

The present study also has several limitations. First, the majority of the apps about food allergies or intolerances found in the app stores had fewer than 1000 reviews and a user star rating <3, indicating low interest by users. Consequently, it was not possible to include most of the apps because we considered a rating of 3 stars as the minimum threshold for app quality. However, it was important for the inclusion criteria to limit the findings to the most reliable and popular apps, as the market includes plenty of dubious apps. Second, apps with only a paid version were excluded from the search. Third, several apps were excluded because of technical problems, such as being unable to open or use the app. Fourth, because this study is not a systematic review of the literature but is a systematic search of app stores, it was not possible to register it in PROSPERO [76]. Finally, despite the increasing attention to apps, the literature about the assessment of app quality is very scarce [77] and not oriented to food allergies and intolerances.

Conclusions

In this systematic search of food allergy or intolerance apps, acceptable MARS quality was identified, although the engagement of food product and restaurant apps should be improved and the included apps should be tested in trials. The critical points identified in this systematic search can help improve the innovativeness and applicability of future food allergy and intolerance apps.

Acknowledgments

This publication was cofunded by the European Regional Development Fund (ERDF) of the European Union within the framework of the ERDF operative programme of Catalonia 2014-2020 aimed at an objective of investment in growth and employment. This publication is framed within the initiative of coordinated PECT TurisTIC en familia, Operation 12: "Healthy Meals". The NFOC-Salut group is a consolidated research group of Generalitat de Catalunya, Spain (2017 SGR522).

This publication received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713679. This publication has been possible with the support of the Universitat Rovira i Virgili (URV) and Banco Santander.

Conflicts of Interest

None declared.

Multimedia Appendix 1
General characteristics of the mobile phone apps included in the review.
[PDF File (Adobe PDF File), 313 KB - mhealth_v8i9e18339_app1.pdf]

Multimedia Appendix 2
Output features of the included apps.
[PDF File (Adobe PDF File), 146 KB - mhealth_v8i9e18339_app2.pdf]

Multimedia Appendix 3
Input features of the included apps.
[PDF File (Adobe PDF File), 125 KB - mhealth_v8i9e18339_app3.pdf]

Multimedia Appendix 4
Mobile App Rating Scale (MARS) and user star ratings of the included apps in mean (SD).
[PDF File (Adobe PDF File), 217 KB - mhealth_v8i9e18339_app4.pdf]

References

1. Montalto M, Santoro L, D'Onofrio F, Curigliano V, Gallo A, Visca D, et al. Adverse reactions to food: allergies and intolerances. Dig Dis 2008;26(2):96-103. [doi: 10.1159/000116766] [Medline: 18431058]
2. Tang R, Wang Z, Ji C, Leung PSC, Woo E, Chang C, et al. Regional Differences in Food Allergies. Clin Rev Allergy Immunol 2019 Aug 5;57(1):98-110. [doi: 10.1007/s12016-018-8725-9] [Medline: 30612248]
3. Turnbull JL, Adams HN, Gorard DA. Review article: the diagnosis and management of food allergy and food intolerances. Aliment Pharmacol Ther 2015 Jan 14;41(1):3-25 [FREE Full text] [doi: 10.1111/apt.12984] [Medline: 25316115]
4. Tuck CJ, Biesiekierski JR, Schmid-Grendelmeier P, Pohl D. Food Intolerances. Nutrients 2019 Jul 22;11(7):1684 [FREE Full text] [doi: 10.3390/nu11071684] [Medline: 31336652]
5. Lomer MCE. Review article: the aetiology, diagnosis, mechanisms and clinical evidence for food intolerance. Aliment Pharmacol Ther 2015 Feb 03;41(3):262-275 [FREE Full text] [doi: 10.1111/apt.13041] [Medline: 25471897]
6. Jones SM, Burks AW, Dupont C. State of the art on food allergen immunotherapy: oral, sublingual, and epicutaneous. J Allergy Clin Immunol 2014 Feb;133(2):318-323. [doi: 10.1016/j.jaci.2013.12.1040] [Medline: 24636471]
7. Licari A, Manti S, Marseglia A, Brambilla I, Votto M, Castagnoli R, et al. Food Allergies: Current and Future Treatments. Medicina (Kaunas) 2019 May 01;55(5):120 [FREE Full text] [doi: 10.3390/medicina55050120] [Medline: 31052434]
8. European Parliament and of the Council of the European Union. Regulation (EU) No 1169/2011 of the European Parliament and of the Council. Official Journal of the European Union. 2011. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32011R1169 [accessed 2020-08-31]
9. Loerbroks A, Tolksdorf SJ, Wagenmann M, Smith H. Food allergy knowledge, attitudes and their determinants among restaurant staff: A cross-sectional study. PLoS One 2019 Apr 24;14(4):e0214625 [FREE Full text] [doi: 10.1371/journal.pone.0214625] [Medline: 31017913]
10. Muraro A, Agache I, Clark A, Sheikh A, Roberts G, Akdis CA, European Academy of Allergy and Clinical Immunology. EAACI food allergy and anaphylaxis guidelines: managing patients with food allergy in the community. Allergy 2014 Aug 18;69(8):1046-1057. [doi: 10.1111/all.12441] [Medline: 24905609]
11. Çelik D, Elçi A, Akçiçek R, Gökçe B, Hürcan P. A safety food consumption mobile system through semantic web technology. 2014 Presented at: 2014 IEEE 38th International Computer Software and Applications Conference Workshops; July 21-25, 2014; Vasteras, Sweden. [doi: 10.1109/compsacw.2014.126]
12. Ross GMS, Bremer MGEG, Nielen MWF. Consumer-friendly food allergen detection: moving towards smartphone-based immunoassays. Anal Bioanal Chem 2018 Sep 26;410(22):5353-5371 [FREE Full text] [doi: 10.1007/s00216-018-0989-7] [Medline: 29582120]
13. Lu C, Hu Y, Xie J, Fu Q, Leigh I, Governor S, et al. The Use of Mobile Health Applications to Improve Patient Experience: Cross-Sectional Study in Chinese Public Hospitals. JMIR Mhealth Uhealth 2018 May 23;6(5):e126 [FREE Full text] [doi: 10.2196/mhealth.9145] [Medline: 29792290]
14. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
15. Oreskovic NM, Huang TT, Moon J. Integrating mHealth and Systems Science: A Combination Approach to Prevent and Treat Chronic Health Conditions. JMIR Mhealth Uhealth 2015 Jun 02;3(2):e62 [FREE Full text] [doi: 10.2196/mhealth.4150] [Medline: 26036753]
16. Silva BMC, Rodrigues JJPC, 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]

17. Yang Q, Van Stee SK. The Comparative Effectiveness of Mobile Phone Interventions in Improving Health Outcomes: Meta-Analytic Review. JMIR Mhealth Uhealth 2019 Apr 03;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]
18. Matricardi PM, Dramburg S, Alvarez-Perea AD, Antolín-Amérigo DS, Apfelbacher C, Atanaskovic-Markovic M, et al. The role of mobile health technologies in allergy care: An EAACI position paper. Allergy 2020 Feb 16;75(2):259-272. [doi: 10.1111/all.13953] [Medline: 31230373]
19. Arens-Volland A, Spassova L, Rösch N. Review of Mobile Health (mHealth) Solutions for Food-related Conditions and Nutritional Risk Factors. 2013 Presented at: eTELEMED 2013: The Fifth International Conference on eHealth, Telemedicine, and Social Medicine; February 24 - March 1, 2013; Nice, France p. 284-289. [doi: 10.1016/b978-0-12-814309-4.00012-4]
20. Clement J. Global digital population as of April 2020. Statista. 2020. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/#:~:text=How [accessed 2020-06-15]
21. Cuervo-Pardo L, Barcena-Blanch MA, Gonzalez-Estrada A, Schroer B. Apps for food allergy: A critical assessment. J Allergy Clin Immunol Pract 2015 Nov;3(6):980-1.e1. [doi: 10.1016/j.jaip.2015.06.011] [Medline: 26246126]
22. Pereira AM, Jácome C, Almeida R, Fonseca JA. How the Smartphone Is Changing Allergy Diagnostics. Curr Allergy Asthma Rep 2018 Oct 25;18(12):69. [doi: 10.1007/s11882-018-0824-4] [Medline: 30361774]
23. Kuehnhausen M, Frost VS. Trusting smartphone Apps? To install or not to install, that is the question. 2013 Presented at: 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA); February 25-28, 2013; San Diego, CA p. 30-37. [doi: 10.1109/cogsima.2013.6523820]
24. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the Quality of Mobile Apps for the Management of Pain: Systematic Search and Evaluation Using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
25. WHO Global Observatory for eHealth. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2020-08-29]
26. US Food and Drug Administration. Policy for Device Software Functions and Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff. 2019 Sep 27. URL: https://www.fda.gov/media/80958/download [accessed 2020-08-29]
27. Generalitat de Catalunya. TIC Salut Social: Pla Mestre de Mobilitat "mHealth.Cat". 2015. URL: https://ticsalutsocial.cat/area/mhealth/ [accessed 2020-06-15]
28. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
29. Dayton SJ. Rethinking Health App Regulation: The Case for Centralized FDA Voluntary Certification of Unregulated Non-Device Mobile Health Apps. IHLR 2013 Dec 31;11(2). [doi: 10.18060/18892]
30. Carpenter D, Grimmer J, Lomazoff E. Approval regulation and endogenous consumer confidence: Theory and analogies to licensing, safety, and financial regulation. Regulation & Governance 2010 Nov 22;4(4):383-407. [doi: 10.1111/j.1748-5991.2010.01091.x]
31. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
32. Aitken M, Gauntlett C. Patient apps for improved healthcare from novelty to mainstream. IMS Institute for Healthcare Informatics. 2013 Oct. URL: http://moodle.univ-lille2.fr/pluginfile.php/215345/mod_resource/content/0/IIHI_Patient_Apps_Report.pdf [accessed 2020-08-29]
33. mHIMSS App Usability Work Group. mHIMSS transforming healthcare. Selecting a Mobile App: Evaluating the Usability of Medical Applications. 2012. URL: http://s3.amazonaws.com/rdcms-himss/files/production/public/HIMSSguidetoappusabilityv1mHIMSS.pdf [accessed 2020-08-29]
34. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular Nutrition-Related Mobile Apps: A Feature Assessment. JMIR Mhealth Uhealth 2016 Aug 01;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
35. Ferrara G, Kim J, Lin S, Hua J, Seto E. A Focused Review of Smartphone Diet-Tracking Apps: Usability, Functionality, Coherence With Behavior Change Theory, and Comparative Validity of Nutrient Intake and Energy Estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
36. Lyzwinski LN, Edirippulige S, Caffery L, Bambling M. Mindful Eating Mobile Health Apps: Review and Appraisal. JMIR Ment Health 2019 Aug 22;6(8):e12820 [FREE Full text] [doi: 10.2196/12820] [Medline: 31441431]
37. Choi YK, Demiris G, Lin S, Iribarren SJ, Landis CA, Thompson HJ, et al. Smartphone Applications to Support Sleep Self-Management: Review and Evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790 [FREE Full text] [doi: 10.5664/jcsm.7396] [Medline: 30353814]
38. Mauch CE, Wycherley TP, Laws RA, Johnson BJ, Bell LK, Golley RK. Mobile Apps to Support Healthy Family Food Provision: Systematic Assessment of Popular, Commercially Available Apps. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11867 [FREE Full text] [doi: 10.2196/11867] [Medline: 30578213]

39. Davis SF, Ellsworth MA, Payne HE, Hall SM, West JH, Nordhagen AL. Health Behavior Theory in Popular Calorie Counting Apps: A Content Analysis. JMIR Mhealth Uhealth 2016 Mar 02;4(1):e19 [FREE Full text] [doi: 10.2196/mhealth.4177] [Medline: 26935898]
40. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
41. Simões P, Silva AG, Amaral J, Queirós A, Rocha NP, Rodrigues M. Features, Behavioral Change Techniques, and Quality of the Most Popular Mobile Apps to Measure Physical Activity: Systematic Search in App Stores. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11281 [FREE Full text] [doi: 10.2196/11281] [Medline: 30368438]
42. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
43. Stoyanov S, Leanne H, Kavanagh D, Tjondronegoro D, Zelenko O, Mani M. MARS training video. Youtube. 2016 Jun 14. URL: https://www.youtube.com/watch?v=25vBwJQIOcE [accessed 2019-08-10]
44. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg 2018 May;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]
45. Boulos MNK, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 May 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
46. Silvennoinen J, Vogel M, Kujala S. Experiencing Visual Usability and Aesthetics in Two Mobile Application Contexts. Journal of Usability Studies 2014 Nov;10(1):46-62 [FREE Full text]
47. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile App for Improved Self-Management of Type 2 Diabetes: Multicenter Pragmatic Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
48. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and Increasing Physical Activity Among Young People: The AIMFIT Pragmatic Randomized Controlled Trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
49. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Dec 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
50. Soon JM. Structural modelling of food allergen knowledge, attitude and practices among consumers in Malaysia. Food Res Int 2018 Sep;111:674-681. [doi: 10.1016/j.foodres.2018.06.001] [Medline: 30007732]
51. Ju S, Park J, Kwak T, Kim K. Attitudes and preferences of consumers toward food allergy labeling practices by diagnosis of food allergies. Nutr Res Pract 2015 Oct;9(5):517-522 [FREE Full text] [doi: 10.4162/nrp.2015.9.5.517] [Medline: 26425282]
52. Begen FM, Barnett J, Payne R, Gowland MH, DunnGalvin A, Lucas JS. Eating out with a food allergy in the UK: Change in the eating out practices of consumers with food allergy following introduction of allergen information legislation. Clin Exp Allergy 2018 Mar;48(3):317-324. [doi: 10.1111/cea.13072] [Medline: 29220107]
53. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. Designing Health Apps to Support Dietetic Professional Practice and Their Patients: Qualitative Results From an International Survey. JMIR Mhealth Uhealth 2017 Mar 31;5(3):e40 [FREE Full text] [doi: 10.2196/mhealth.6945] [Medline: 28363882]
54. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User Preferences and Design Recommendations for an mHealth App to Promote Cystic Fibrosis Self-Management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
55. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth Design Space: A Mixed-Methods Analysis of Smokers' Perspectives. JMIR Mhealth Uhealth 2016 Aug 05;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
56. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
57. Wilson H, Stoyanov SR, Gandabhai S, Baldwin A. The Quality and Accuracy of Mobile Apps to Prevent Driving After Drinking Alcohol. JMIR Mhealth Uhealth 2016 Aug 08;4(3):e98 [FREE Full text] [doi: 10.2196/mhealth.5961] [Medline: 27502956]
58. Woldaregay A, Issom DZ, Henriksen AL, Marttila HR, Mikalsen M, Pfuhl G, et al. Motivational Factors for User Engagement with mHealth Apps. Stud Health Technol Inform 2018;249:151-157. [Medline: 29866972]
59. Bidargaddi N, Pituch T, Maaieh H, Short C, Strecher V. Predicting which type of push notification content motivates users to engage in a self-monitoring app. Prev Med Rep 2018 Oct;11:267-273 [FREE Full text] [doi: 10.1016/j.pmedr.2018.07.004] [Medline: 30109172]

60. Marston HR, Hall AK. Gamification: Applications for health promotion and health information technology engagement. In: Novák D, Brendryen H, Tulu B, editors. Handbook of Research on Holistic Perspectives in Gamification for Clinical Practice. Hershey, PA: IGI Global; 2015:78-104.
61. Ignacia SN, Wiastuti RD, Lemy DM. Restaurant Mobile Application towards Purchase Intention. IJAST 2018 Aug 31;117:113-128. [doi: 10.14257/ijast.2018.117.10]
62. Yu Y, Luo M, Zhu D. The Effect of Quality Attributes on Visiting Consumers' Patronage Intentions of Green Restaurants. Sustainability 2018 Apr 15;10(4):1187. [doi: 10.3390/su10041187]
63. Jeminiwa RN, Hohmann NS, Fox BI. Developing a Theoretical Framework for Evaluating the Quality of mHealth Apps for Adolescent Users: A Systematic Review. J Pediatr Pharmacol Ther 2019;24(4):254-269 [FREE Full text] [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
64. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Mar;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
65. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J. Awareness and Use of mHealth Apps: A Study from England. Pharmacy (Basel) 2017 Jul 14;5(2) [FREE Full text] [doi: 10.3390/pharmacy5020033] [Medline: 28970445]
66. King R, Knibb RC, Hourihane JO. Impact of peanut allergy on quality of life, stress and anxiety in the family. Allergy 2009 Mar;64(3):461-468. [doi: 10.1111/j.1398-9995.2008.01843.x] [Medline: 19076542]
67. Zarkadas M, Dubois S, MacIsaac K, Cantin I, Rashid M, Roberts KC, et al. Living with coeliac disease and a gluten-free diet: a Canadian perspective. J Hum Nutr Diet 2013 Mar;26(1):10-23. [doi: 10.1111/j.1365-277X.2012.01288.x] [Medline: 23157646]
68. Cummings A, Knibb RC, King RM, Lucas JS. The psychosocial impact of food allergy and food hypersensitivity in children, adolescents and their families: a review. Allergy 2010 Aug;65(8):933-945. [doi: 10.1111/j.1398-9995.2010.02342.x] [Medline: 20180792]
69. Muñoz-Furlong A. Daily coping strategies for patients and their families. Pediatrics 2003 Jul;111(6 Pt 3):1654-1661. [Medline: 12777606]
70. Wilhide Iii CC, Peeples MM, Anthony Kouyaté RC. Evidence-Based mHealth Chronic Disease Mobile App Intervention Design: Development of a Framework. JMIR Res Protoc 2016 Mar 16;5(1):e25 [FREE Full text] [doi: 10.2196/resprot.4838] [Medline: 26883135]
71. Davis A, Ellis R. A quasi-experimental investigation of college students' ratings of two physical activity mobile apps with varied behavior change technique quantity. Digit Health 2019;5:2055207619891347 [FREE Full text] [doi: 10.1177/2055207619891347] [Medline: 31827878]
72. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016 Jul 07;16:83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
73. Zhu H, Xiong H, Ge Y, Chen E. Discovery of Ranking Fraud for Mobile Apps. IEEE Trans. Knowl. Data Eng 2015 Jan;27(1):74-87. [doi: 10.1109/TKDE.2014.2320733]
74. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
75. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for People With Rheumatoid Arthritis to Monitor Their Disease Activity: A Review of Apps for Best Practice and Quality. JMIR Mhealth Uhealth 2017 Mar 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]
76. PROSPERO International Prospective register of systematic reviews. National Institute for Health Research. URL: https://www.crd.york.ac.uk/PROSPERO/ [accessed 2020-08-29]
77. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]

Abbreviations

IMS: Intercontinental Medical Statistics
MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by G Eysenbach; submitted 21.02.20; peer-reviewed by M Khalemsky, T Muto, C Reis, E Chioma, M K.; comments to author 02.04.20; revised version received 27.05.20; accepted 25.06.20; published 16.09.20.

Please cite as:
Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R
Mobile Phone Apps for Food Allergies or Intolerances in App Stores: Systematic Search and Quality Assessment Using the Mobile App Rating Scale (MARS)
JMIR Mhealth Uhealth 2020;8(9):e18339
URL: http://mhealth.jmir.org/2020/9/e18339/
doi:10.2196/18339
PMID:32936078

Original Paper

Development and Evaluation of a Tailored Mobile Health Intervention to Improve Medication Adherence in Black Patients With Uncontrolled Hypertension and Type 2 Diabetes: Pilot Randomized Feasibility Trial

Antoinette Schoenthaler[1*], EdD; Michelle Leon[2*], BA; Mark Butler[3*], PhD; Karsten Steinhaeuser[4*], MS, PhD; William Wardzinski[4*], BS

[1]Department of Population Health, NYU school of Medicine, Center for Healthful Behavior Change, NYU Langone Health, New York, NY, United States

[2]Department of Clinical Psychology, Fordham University, New York, NY, United States

[3]Center for Personalized Health, Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, United States

[4]Radiant Point Technologies, Inc., Honolulu, HI, United States

[*]all authors contributed equally

Corresponding Author:
Antoinette Schoenthaler, EdD
Department of Population Health
NYU School of Medicine, Center for Healthful Behavior Change
NYU Langone Health
180 Madison Ave
7th floor
New York, NY, 10016
United States
Phone: 1 6465013434
Email: antoinette.schoenthaler@nyumc.org

Abstract

Background: Research has underscored the need to develop socioculturally tailored interventions to improve adherence behaviors in minority patients with hypertension (HTN) and type 2 diabetes (T2D). Novel mobile health (mHealth) approaches are potential methods for delivering tailored interventions to minority patients with increased cardiovascular risk.

Objective: This study aims to develop and evaluate the acceptability and preliminary efficacy of a tailored mHealth adherence intervention versus attention control (AC) on medication adherence, systolic blood pressure (SBP), diastolic blood pressure (DBP), and hemoglobin A_{1c} (HbA_{1c}) at 3 months in 42 Black patients with uncontrolled HTN and/or T2D who were initially nonadherent to their medications.

Methods: This was a two-phase pilot study consisting of a formative phase and a clinical efficacy phase. The formative phase consisted of qualitative interviews with 10 members of the target patient population (7/10, 70% female; mean age 65.8 years, SD 5.6) to tailor the intervention based on the Information-Motivation-Behavioral skills model of adherence. The clinical efficacy phase consisted of a 3-month pilot randomized controlled trial to evaluate the tailored mHealth intervention versus an AC. The tablet-delivered intervention included a tailoring survey, an individualized adherence profile, and a personalized list of interactive adherence-promoting modules, whereas AC included the tailoring survey and health education videos delivered on the tablet. Acceptability was assessed through semistructured exit interviews. Medication adherence was assessed using the 8-item Morisky Medication Adherence Scale, whereas blood pressure and HbA_{1c} were assessed using automated devices.

Results: In phase 1, thematic analysis of the semistructured interviews revealed the following 5 major barriers to adherence: disruptions in daily routine, forgetfulness, concerns about adverse effects, preference for natural remedies, and burdens of medication taking. Patients recommended the inclusion of modules that address improving patient-provider communication, peer vignettes, and stress reduction strategies to facilitate adherence. A total of 42 Black patients (23/42, 55% male; mean age 57.6 years, SD 11.1) participated in the clinical efficacy pilot trial. At 3 months, both groups showed significant improvements in adherence (mean 1.35, SD 1.60; $P<.001$) and SBP (−4.76 mm Hg; $P=.04$) with no between-group differences ($P=.50$ and $P=.10$).

The decreases in DBP and HbA_{1c} over time were nonsignificant (−1.97 mm Hg; *P*=.20; and −0.2%; *P*=.45, respectively). Patients reported high acceptability of the intervention for improving their adherence.

Conclusions: This pilot study demonstrated preliminary evidence on the acceptability of a tailored mHealth adherence intervention among a sample of Black patients with uncontrolled HTN and T2D who were initially nonadherent to their medications. Future research should explore whether repeated opportunities to use the mHealth intervention would result in improvements in behavioral and clinical outcomes over time. Modifications to the intervention as a result of the pilot study should guide future efforts.

Trial Registration: ClinicalTrials.gov NCT01643473; http://clinicaltrials.gov/ct2/show/ NCT01643473

(JMIR Mhealth Uhealth 2020;8(9):e17135) doi:10.2196/17135

KEYWORDS

mHealth; medication adherence; hypertension; type 2 diabetes; African Americans

Introduction

Background

Despite advances in treatments for hypertension (HTN) and type 2 diabetes (T2D), Black patients continue to experience disproportionately lower rates of blood pressure (BP) and glycemic control than those observed in White patients [1,2]. Poor medication adherence among Black patients may explain the disproportionately lower rates of BP and glucose control in this patient population than in White patients [3,4]. Compared with their White counterparts, Black patients with HTN and T2D have been shown to be 1.81 to 4.30 times less likely to adhere to their medication regimen [4-6]. Given that a sufficiently high level of adherence is key for achieving adequate disease control, it follows that successful approaches to reducing the racial gap in cardiovascular-related mortality that exist between Black and White patients must take into consideration the factors driving poor medication adherence in Black population.

Despite a wealth of research dedicated to understanding adherence behaviors in patients with T2D and HTN [3,4], trials designed to improve adherence in minority patients have shown limited effectiveness [7]. Several investigators have called attention to the need for tailored interventions to improve medication adherence in minority patients [7,8], with *tailoring* referring both to cultural tailoring (eg, medication beliefs) and adapting the intervention to match patients' needs and preferences. Increasingly, mobile health (mHealth) technologies, such as mobile phones, tablets, and other personal digital assistants, are being used as efficient and acceptable methods for delivering tailored interventions to patients with increased cardiovascular risk. Several systematic reviews have documented the short-term benefits of mHealth interventions for improving medication adherence in patients with HTN or T2D [9,10]. However, of the mHealth interventions that have aimed to improve medication adherence, only 6 have been conducted in high-risk minority populations with HTN or T2D, all of which used text messages as their primary method of intervention delivery [11-16]. Although text messaging offers several advantages for improving adherence behaviors (eg, sending reminder prompts in real time), several shortcomings have also been noted. For example, with text messaging, only brief educational, motivational, and/or behavioral content within a limited number of characters can be provided. As a result, they can lack depth by not covering all the necessary content and require individuals to access links for supplementary materials (eg, through videos) [17]. Although personalization is possible with text messaging, qualitative feedback from studies also note that content can become repetitive and predictable, leading to message fatigue and disengagement from the intervention [18,19]. Reading and responding to text messages also require a level of visual acuity and dexterity that may be challenging for people who experience any motor and/or visual impairments. This is especially true for patients with uncontrolled T2D and/or HTN who may experience retinopathy as a result of their disease.

Owing to their portability and ease of use, tablets are increasingly being used as an acceptable digital platform to deliver interventions across all age and racial and ethnic groups [20,21]. Tablet devices offer several advantages over text messages. This includes an adjustable font or icon size, a touchscreen, the ability to integrate video and auditory features into the intervention, which may be better suited for individuals with chronic disease and people with lower levels of lower health literacy, and the ability to create a more interactive and hands-on learning environment.

Objectives

Consequently, the development of tailored mHealth interventions that use alternative digital platforms to improve medication adherence in Black patients is needed to address the marked racial disparities in BP and glycemic control. In this paper, we report the development and evaluation of an interactive tablet-delivered intervention that was socioculturally tailored for Black patients with uncontrolled HTN and/or T2D who were initially nonadherent to their medications.

Methods

Design Overview

We conducted a two-phase feasibility study [22] that included a formative phase and a clinical efficacy phase. The formative phase consisted of qualitative interviews to tailor the intervention to the needs and preferences of the target population. For the clinical efficacy phase, we evaluated the acceptability and preliminary efficacy of the tailored mHealth intervention versus an attention control (AC) condition on changes in medication adherence, systolic BP (SBP), diastolic BP (DBP), and hemoglobin A_{1c} (HbA_{1c}) at 3 months

(exploratory outcomes) among a sample of 42 Black patients with uncontrolled HTN and/or T2D who were initially nonadherent to their medications in a pilot randomized controlled trial (RCT). We hypothesized that the mHealth intervention would be acceptable and result in better medication adherence and a greater reduction in SBP, DBP, and HbA_{1c} at 3 months compared with the AC group.

Setting and Participants

This study was conducted at a safety net primary care clinic in New York City, which serves a predominately diverse, low-income urban patient population. Eligibility criteria for both phases of the study included patients who self-identified as Black or African American, received care at the primary care clinic, had uncontrolled HTN defined as BP>140/90 mm Hg (or BP>130/80 mm Hg for those with diabetes or kidney disease) and/or uncontrolled T2D defined as HbA_{1c}>7% on at least two visits in the past year and at least one cardiovascular risk factor (eg, hyperlipidemia or obesity), had been prescribed at least one antihypertensive or oral antidiabetic medication and were nonadherent to their medication at screening (as described in the following paragraph), were at least 18 years old, were fluent in English, and did not have significant psychiatric comorbidity. The Institutional Review Board of New York University approved the study.

Potentially eligible patients for both phases were identified through a review of the electronic medical records, after which letters, signed by the physician, were sent to patients inviting them to participate in the study. A trained research assistant (RA) completed all screening, consent, and data collection procedures. During the screening procedures, medication adherence was assessed using the validated 8-item Morisky Medication Adherence Scale (MMAS-8) [23-25]. The first 7 items require a yes or no response, and the final item uses a 5-point scale (*never/rarely* to *all the time*). Total MMAS-8 scores range from 0 to 8, with a score of <6 indicating nonadherence [23]. Only patients with a score <6 were eligible to participate in the study. After obtaining patients' written informed consent and completion of baseline measures, the statistician randomized eligible patients in a ratio of 1:1 to either the intervention or AC condition using block randomization, with the investigators blinded to the permutation. Following Consolidated Standards of Reporting Trials guidelines [26], the randomization sequence was kept in a secure electronic file that only restricted staff could access. Given the nature of the intervention, patients could not be blinded to the group assignment [27]. However, we used automated BP and HbA_{1c} devices to lower the likelihood that the RA could influence the clinical outcomes.

Formative Phase: Development of the mHealth Intervention

The Information-Motivation-Behavioral skills (IMB) model of adherence is the theoretical framework underlying the intervention [28,29]. This model views the interrelations between adherence-related information (eg, how medications work), motivation (eg, attitudes or beliefs), and behavioral skills (eg, self-efficacy to take medications) as the fundamental determinants of behavior. A trained qualitative researcher (AS) conducted semistructured interviews in a dedicated, private space with a convenience sample of 10 Black patients before the initiation of the trial. Each interview was approximately 30 min in duration, audiotaped, and professionally transcribed.

The constructs of the IMB model were used to develop the interview questions and probes that were asked in the semistructured interviews. Specifically, questions targeted the most salient informational (eg, how medications work, side effects), motivational (eg, social support, beliefs), and behavioral (eg, self-efficacy, ability to administer medications) barriers and facilitators that may affect adherence behaviors. Interview questions on the most salient barriers to adherence were constructed using the IMB survey items and from the existing adherence literature, including the authors' research in this patient population [28,30,31]. Sample interview questions included, "What, if any, reasons did your doctor give you about why s/he felt that you needed to take blood pressure/diabetes medicines? (information)," "What, if any, concerns do you have about the medications you are taking for your HTN/T2D? (motivation)," and "Tell me about situations or times that make it more difficult to take your HTN/T2D medications (e.g., when traveling, at work, when costs are too high)? (behavioral skills)." To better fit the needs, beliefs, and experiences of Black patients with HTN and/or T2D, interview questions were also used to identify the sociocultural barriers and facilitators to adherence that were not captured in the survey [32,33]. An example question about sociocultural factors was "Tell me about situations when you have used home remedies to improve your blood pressure/diabetes? What specific home remedies do you take?"

Description of the mHealth Intervention Group

The mHealth intervention was built by Radiant Point Technologies using Microsoft's Models, Views, and Controllers Entity Framework as the development environment for the intervention. The intervention consists of an administrative interface for creating user accounts and exporting data and a patient portal for entering information (eg, user profile, questionnaires) and completing activity modules. The fully automated intervention consisted of 3 main parts: (1) a *tailoring survey* based on the IMB Adherence Questionnaire [34], (2) an *individualized adherence profile*, and (3) a personalized list of *interactive adherence-promoting modules* that were matched to the barriers outlined on the adherence profile.

Figure 1 shows the flow of the intervention for participants randomized to the mHealth intervention arm. Once randomized to the intervention group, the RA escorted patients to a private room and provided them with a tablet and instructed them on how to begin the program via a password-protected portal. Patients then completed the IMB tailoring survey, and their responses were immediately scored with an automated algorithm that calculated their 2 most salient adherence barriers (see *Study Measures* section for a description of the scoring algorithm). These data were used to create an individualized adherence profile for each patient in the intervention group, which described the 2 most salient adherence barriers identified by the survey and displayed a personalized list of up to 6

adherence-promoting strategies that were matched to the adherence barriers (Table 1). Patients then had the opportunity to select and work through any (or all) of the strategies that they felt were of the greatest importance and utility for improving their medication adherence. At the end of the program, patients developed an adherence action plan using the principles of specific, measurable, achievable, relevant, and time-bound goal setting. As a feasibility study, patients only interacted with the intervention once at the time of their baseline visit.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

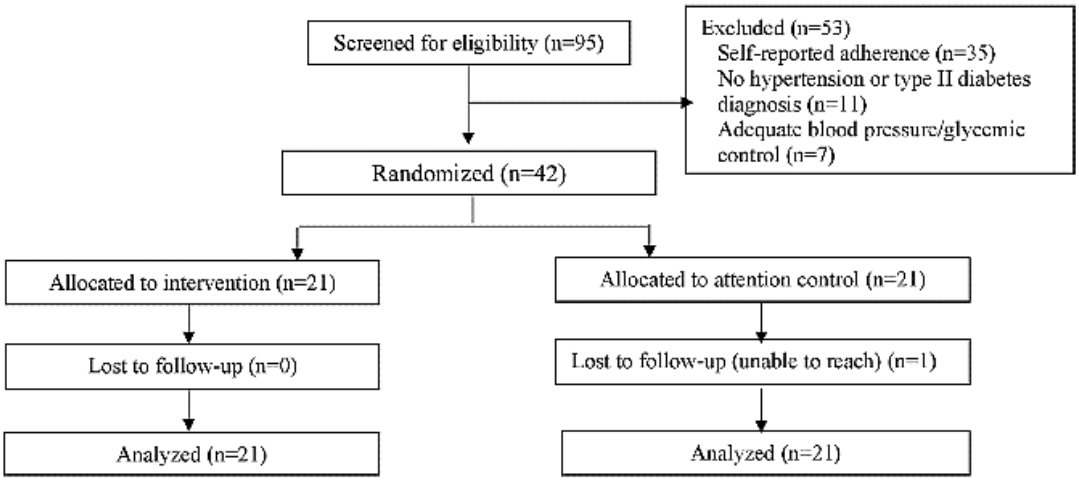


Table 1. Intervention Information-Motivation-Behavioral constructs and matched adherence-promoting strategies.

Information-Motivation-Behavioral construct and modules	Example intervention strategies
Information (knowledge about HTN[a] and medication regimen; side effects and drug interactions)	
The City map	• Interactive map that allows patient to choose different buildings (eg, hospital, community clinic) that describe local and national services for prescription assistance
Pharmacist corner	• Interactive prescription label that allows patients to select areas on the label to learn more about what the information means and why it is important
Motivation (individual and social) (beliefs or attitudes [ie, illness perceptions, concerns]; social norms or influence; perceived efficacy; depression or stress)	
Helping hands	• Narratives by Black patients that discuss the importance of taking medications in context of their life values (ie, religious beliefs, family coherence), strategies to talk to their doctor about medications, and how to develop routines to take medications every day • Positive voice videos that allow patients to hear about other Black patients' experiences with HTN and type 2 diabetes and how they overcame challenges to taking medications, as prescribed
Relaxation station	• Interactive body map that allows patients to learn how common stressors affect their health • Guided relaxation activity • Discussion on the use of prayer and affirmations to combat the negative effects of stress
Doc-Talk	• Question building section that allows patients to develop a list of questions they would like to ask at their next visit • Tip sheet on how to express concerns about and goals for medications to providers
Behavioral skills (habituation and vigilance; routine; ability [subjective and objective])	
Myth busters	• Interactive game to increase disease- and regimen-specific knowledge as well as address misconceptions or beliefs about medications through a true or false quiz
Habit formation	• Development of if-then statements that help patients develop habits to take medications even when their routines are disrupted
Goal setting	• Develop specific, measurable, achievable, relevant, time-bound goals for adherence • Celebrate success that allows patients with perfect adherence to create a reward certificate

[a]HTN: hypertension.

Description of the AC Group

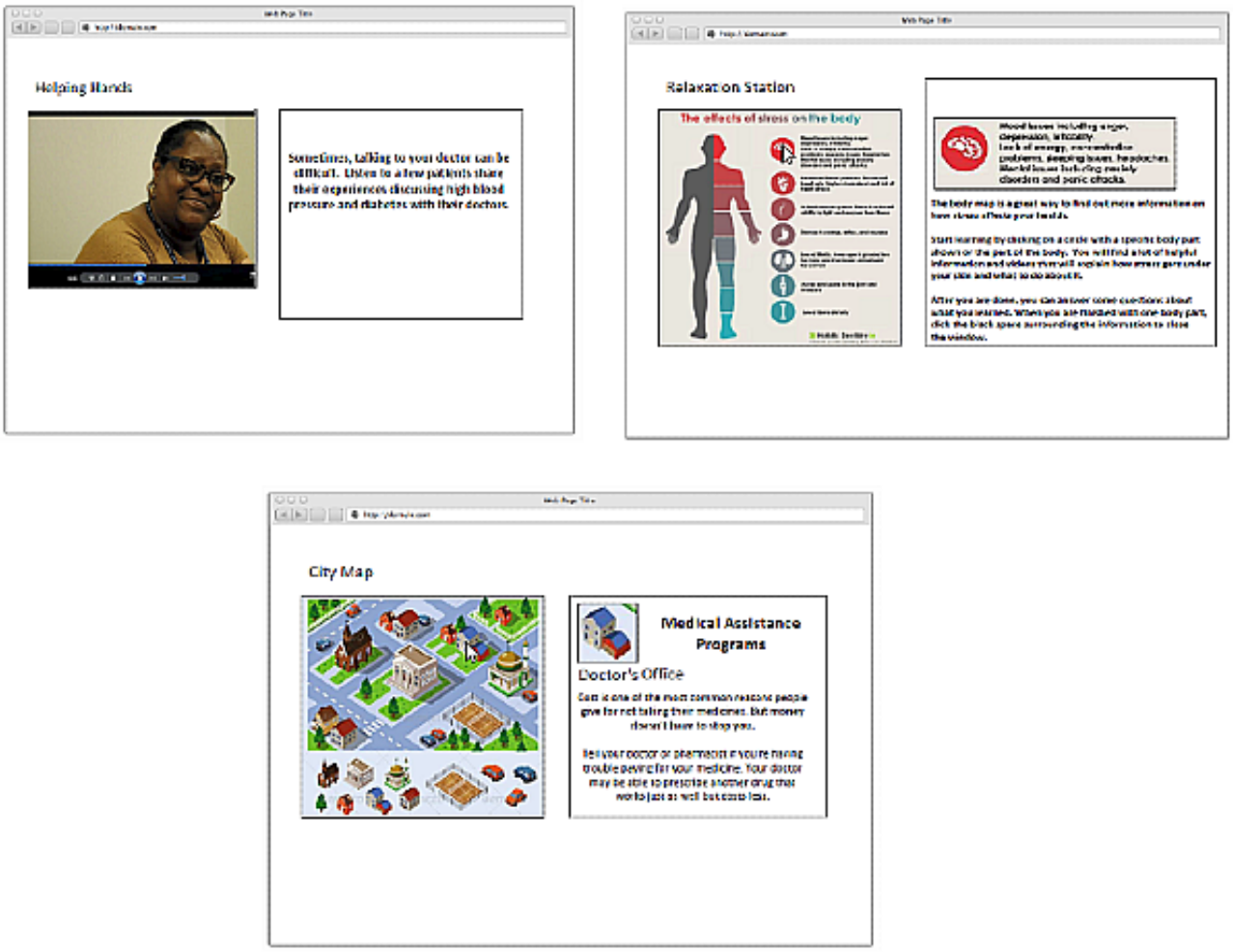
To control for attention and novelty of the technology, patients randomized to the AC group completed the introductory tailoring survey on the same platform as the patients in the intervention group; however, they did not receive the results displaying their 2 most salient adherence barriers. After the completion of the tailoring survey, the program directed patients to a menu of health education modules on topics unrelated to medication adherence. The duration of the modules was the same as that of the intervention; the modules included basic information derived from the resources published by national organizations (eg, National Cancer Institute) on areas such as the cause and consequences of the disease, associated risk factors, and lifestyle changes.

Study Measures

Acceptability

The acceptability of the intervention was assessed through exit interviews conducted with patients in the intervention group at the 3-month visit. Questions inquired about the perceived ease of use of the tablet-delivered intervention, usefulness of the intervention to address patients' adherence barriers, relevance of the content, satisfaction with the different intervention modules (Figure 2), and recommendations for improvement.

Figure 2. Screenshots of intervention modules.



Preliminary Efficacy

The preliminary efficacy of the intervention was assessed at baseline and 3 months using behavioral and clinical outcome measures of medication adherence, reduction in SBP and DBP, and reduction in HbA_{1c}. A 3-month follow-up was chosen to mimic clinical practice for measuring changes in BP and HbA_{1c} in patients with uncontrolled HTN and T2D, respectively.

Medication Adherence

Medication adherence was assessed using the well-validated MMAS-8 score (α=.83) [23]. The MMAS-8 has a reported sensitivity of 93% and specificity of 53% in detecting nonadherence when compared with prescription refill data [23]. Higher scores on the MMAS-8 have been associated with higher rates of uncontrolled BP and poor glycemic control among adults with HTN and T2D, respectively [23].

BP

BP was assessed using validated automated WatchBP monitors (Microlife) at all study visits, following the American Heart Association guidelines [35]. The average of 3 SBP and DBP readings was used as the measurement for each study visit.

HbA_{1c}

HbA_{1c} was assessed using a blood sample drawn via finger-stick and analyzed using a validated point-of-care device (Afinion AS100 Analyzer) that provides HbA_{1c} results in 3 min.

IMB Tailoring Survey

Constructs of the IMB model in the tailoring survey were assessed using a modified version of the empirically validated IMB Adherence Questionnaire [36]. The scale was originally designed to measure the barriers and facilitators of adherence among patients who were HIV positive in clinical care. The 33-item questionnaire comprises 3 subscales, which quantify patients' adherence-related informational (9 items), motivational (10 items), and/or behavioral skills (14 items) and strengths and weaknesses. Responses were given on a 5-point Likert scale, ranging from *I strongly* disagree to *I strongly agree.*

The barriers identified in the IMB survey were used to create individualized adherence profiles that drove the intervention content. Responses given in a *critical range* for each subscale of the questionnaire (ie, response of strongly agree) reflected a significant deficit in the adherence behavior. Automated decision rules and algorithms were used to synthesize the data and generate an individualized adherence profile that summarized the patient's barriers to medication adherence from greatest to least (scores range from 0% [no problem] to 100% [significant

problem]). The most salient barriers for the adherence profile were those with the top 2 relatively higher scores than those with other barriers.

Other Assessments

The RA abstracted clinical data from patients' electronic health record at the initial screening visit and 3-month visit, including duration of HTN and T2D, total number and classes of antihypertensive and oral antidiabetic medications, and comorbid conditions. Data on patient sociodemographics, including age, gender, household income, education level, employment status, and health insurance status were collected from patients at baseline.

Health Literacy

Health literacy was also assessed at baseline using the 36-item short-form Test of Functional Health Literacy in Adults (s-TOFHLA) [37]. The s-TOFHLA is a reading comprehension test that has been linked to poorer health outcomes in racial and ethnic minority populations [37-39]. Total scores of 0 to 16 indicate inadequate health literacy, 17 to 22 indicate marginal health literacy, and 23 to 36 indicate adequate health literacy. Health literacy was included as a covariate in all analytic models.

Analysis

Sample size estimates for the formative phase were based on best practices for maximizing the information power of qualitative research, which recommends beginning with 8 to 10 participants and adding to the sample, as needed [40]. All interviews were audiotaped and transcribed verbatim. Two members of the study team trained in qualitative methods conducted the analysis of the audiotaped interview data. The transcripts of the interviews were uploaded to the Atlas.ti program to facilitate coding and analysis. The transcripts were individually reviewed and analyzed using the grounded theory constant comparison method [41,42]. Specifically, transcripts were coded line by line using open coding (comparing and categorizing data to generate concepts), axial coding (reorganizing data into categories based on relationships within and between these categories), and selective coding (identifying and describing the central themes to generate a conceptual framework) according to facilitators and barriers to medication taking (eg, side effects, cost, forgetfulness, quality of life). Once the transcripts were independently coded, the research team met to discuss the coding and resolve any discrepancies.

As a pilot trial, our sample size estimates for the RCT were exploratory and intended to generate pilot data to calculate the effect sizes needed for a larger trial. On the basis of meta-analyses of adherence interventions, the sample size was calculated using a moderate change in adherence (0.49 between-group difference) [43] as the effect size, power of .80, and significance level of α of .05. This suggested a sample size of 40 patients (20 per group).

Independent *t* tests and chi-square statistics were used to determine if there were any significant differences between consenting participants who dropped out of the study versus completers on any sociodemographic or clinical variables. To assess acceptability, exit interviews were analyzed using grounded theory methods, as described in the *Formative Phase* section. The analyses of the RCT outcomes were performed using an intent-to-treat design. Analysis of covariance models were used to analyze continuous medication adherence, BP, and HbA_{1c} outcomes measured at baseline and 3 months while controlling for baseline values of each outcome measure in their respective models. The outcomes were modeled as functions of time, treatment, and time-by-treatment interaction. Missing data were handled by estimating model parameters for each individual using maximum likelihood estimation based on the available data.

Results

Formative Phase

We invited 13 patients with HTN and/or T2D (4 men and 9 women) to participate in the interviews, of which 3 declined to participate, leaving a total of 10 patients. The reasons for declining participation included being too busy and not being interested in participating in research. Of the 10 patients who agreed to participate, 70% were female, and the mean age was 65.8 years (SD 5.6). The participants varied in their use of technology. Overall, 30% (3/10) of participants exclusively used mobile phones for the primary purposes of talking with family and friends and setting alarms and alerts. Half of the participants used both tablets and mobile devices most commonly to communicate with others and play games. A minority of participants (2/10, 20%) used their devices to track their health (eg, to track medication taking or doctor's appointments).

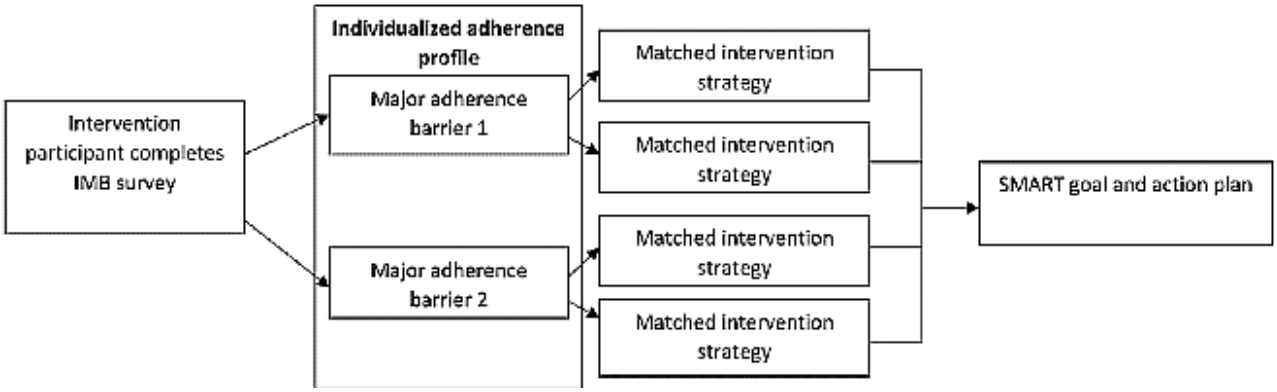
On the basis of qualitative feedback from the interview participants, we made several changes to the wording of the IMB survey to reflect a sample of patients with HTN hypertensive and T2D. For example, the question, "I know how my HIV medications interact with alcohol and street drugs" was revised to state, "I know how my [high blood pressure/diabetes] medications interact with other over the counter medications like cough and cold medicines."

The analysis of the interviews revealed 5 major barriers to adherence: (1) disruptions in daily routines, (2) forgetfulness, (3) concerns about adverse effects, (4) preference for natural remedies, and (5) burdens of medication taking. Specifically, interviewees commented on the challenges of remembering to take their medications when rushing in the morning, traveling, attending appointments, or experiencing other disruptions to their daily routine. Several interviewees also expressed concerns about the side effects of the medications and the potential long-term harm that they may cause to their body. These fears sometimes caused interviewees to "take breaks" from their medications to "let their bodies heal." They also preferred taking natural remedies to treat their HTN and T2D because "all medications, to some degree, are toxic and not from the Earth." Finally, despite acknowledging the need for medications, all the interviewees felt that their life is *limited* because of the medications. For example, one interviewee commented that feeling dependent on insulin constrains their ability "to travel, be active, and just have fun." The short- and long-term concerns

about medications expressed by the participants in the interviews (ie, themes 3-5) were noted as key sociocultural beliefs to address in the intervention to improve adherence behaviors in Black patients.

In total, 3 strategies to promote adherence also emerged from the interviews: collaborative patient-provider communication, use of peer vignettes, and stress reduction techniques. Overall, most interviewees trusted their providers but felt that they lacked the skills to confidently speak with him or her about their medication concerns. They recommended including interactive modules that allowed them to prepare a list of questions they could ask their provider to help facilitate this discussion. Peer vignettes about common challenges others face in managing their HTN and T2D was also a commonly discussed strategy to improve medication adherence. Interviewees recommended including videos featuring peers who share their experiences overcoming challenges to make healthy lifestyle changes to improve their HTN and T2D and tips for integrating medication taking into their daily routine. Finally, all interviewees commented on the importance of including modules that explain how stress affects health and learning stress reduction techniques to improve their emotional and physical well-being. On the basis of these themes, we developed a list of evidence-based adherence-promoting strategies that could help patients address the aforementioned barriers (Table 1 and Figure 3).

Figure 3. Intervention flow. SMART: specific, measurable, achievable, relevant, time-bound.



Each strategy was socioculturally tailored for Black patients based on feedback from the interviews, the research teams' expertise, and the existing literature [7,30,43-47]. For example, the Myth Busters module addressed inaccurate beliefs about medications by including true or false questions about the use of herbal remedies to treat HTN and T2D, the effectiveness of generic medications, and the long-term safety of medications. The Helping Hands video included video testimonials by Black patients who participated in the formative phase interviews and spoke about how they overcame difficulties by asking their provider questions about their medications and how their values supported their decision to make healthy lifestyle changes and take medications, as prescribed. Finally, the Relaxation Station discussed the use of prayer and guided meditation as strategies to lower the negative effects of stress on the body.

Clinical Efficacy Phase

From July 2016 to January 2018, we screened 95 patients for eligibility. Of these patients, we excluded 53 because they did not have an HTN and/or T2D diagnosis (n=11), their condition was under control (n=7), or they reported being adherent to their medications (n=35). Thus, 42 patients participated in the pilot RCT, of which 21 were randomized to each arm (Figure 1). Overall, 42 participants completed the 3-month visit; one patient from the AC group was lost to follow-up. There were no differences between patients who completed the trial and those who were lost to follow-up in terms of any sociodemographic or clinical characteristics (P>.05).

The mean age was 57.6 years (SD 11.1), 54.8% were men, 86% (36/42) had an income of ≤US $40,000 per year, and 55% (23/42) had a high school education (Table 2). Approximately half of the patients (19/42, 45%) had Medicaid. On average, patients were prescribed 3.2 (SD 1.8) antihypertensive and oral diabetic medications. In addition, 95% (40/42) of the patients had adequate health literacy. Patients in the intervention group were significantly more likely to be unemployed than patients in the AC group (20/42, 95% vs 9/42, 43%; P<.001); thus, employment status was entered as a covariate in all analyses.

For both patients with HTN and T2D, negative attitudes or beliefs about medications (motivation-attitude) and greater personal concerns about taking medications (motivation-personal) were the most salient barriers to medication adherence identified by the IMB tailoring survey at baseline (see Multimedia Appendix 1 for data on all the barriers). On the basis of these barriers, the most frequently matched intervention strategies were the *Helping Hands* vignettes (9/21, 43%), the *Myth Busters* game (5/21, 24%), and the *Doc-Talk* activity (3/21, 14%; Table 1).

Table 2. Baseline sociodemographic characteristics for all patients and by study group.

Characteristics	All patients (N=42)	Intervention patients (n=21)	Attention control patients (n=21)
Age (years); range: 36-82, mean (SD)	57.6 (11.1)	59.7 (10.7)	54.5 (11.3)
Gender (male), n (%)	23 (54.8)	11 (52.4)	12 (57.1)
Marital status, n (%)			
Single	11 (26.2)	5 (23.8)	6 (28.6)
Married	8 (19.0)	3 (14.3)	5 (23.8)
Divorced or separated	16 (38.1)	9 (42.9)	7 (33.3)
Widowed	7 (16.7)	4 (19.0)	3 (14.3)
Education, n (%)			
Less than high school	5 (11.9)	2 (9.5)	3 (14.3)
High school or technical school	23 (54.8)	13 (61.9)	10 (47.6)
Some college	6 (14.3)	3 (14.3)	3 (14.3)
College and above	8 (18.8)	3 (14.3)	5 (23.8)
Unemployed	29 (69.0)	20 (95.2)	9 (42.9)
Income (US$), n (%)			
<20,000	26 (62.8)	13 (61.9)	13 (61.9)
20,000-40,000	10 (23.2)	6 (28.5)	4 (19.0)
>40,000	6 (14.0)	2 (9.5)	4 (19.0)
Insurance, n (%)			
Private	8 (19.0)	3 (14.3)	5 (23.8)
Medicare without Medicaid	7 (16.7)	4 (19.0)	3 (14.3)
Medicaid only or with Medicare	19 (45.3)	10 (47.7)	9 (42.9)
None	8 (19.0)	4 (19.0)	4 (19.0)
Health literacy, n (%)			
Inadequate	1 (2.3)	0 (0)	2 (9.5)
Marginal	1 (2.3)	0 (0)	1 (4.8)
Adequate	40 (95.4)	21 (100)	18 (85.7)
Diabetes, n (%)	31 (72.1)	17 (77.3)	14 (66.7)
Stroke, n (%)	6 (14)	3 (14.3)	3 (14.3)
Kidney disease, n (%)	2 (4.7)	2 (9.5)	0 (0)
Number of antihypertensive or oral diabetic medications, mean (SD)	3.20 (1.8)	3.65 (1.8)	2.75 (1.7)

Acceptability of Intervention

Overall, 92% (19/21) of the patients in the intervention group agreed that the mHealth intervention could be an effective tool to help patients take their medications. Specifically, they benefited by learning the importance of taking their medications (17/21, 82%), learning how to speak to their doctor about medication concerns (14/21, 67%), and developing new habits to take medications regularly or make healthy lifestyle changes (19/21, 92%). Most (19/21, 92%) patients rated the vignettes as the best intervention strategy. All patients (21/21, 100%) felt that the intervention was designed for someone like them. The patients suggested the following modifications: (1) shortening the tailoring survey so more time can be spent using the strategies (9/21, 43%), (2) share the results with their doctor to help stimulate conversations about challenges to adherence during the clinic visit (17/21, 82%), and (3) improve tablet button size and sensitivity and audio quality (8/21, 40%). Finally, one-third of patients (7/21, 33%) recommended adding a health educator in addition to the mHealth intervention because they felt it would be beneficial to also discuss questions with a person.

Effect of the Intervention on Medication Adherence

The mean self-reported adherence for the intervention and AC groups at baseline was 4.4 (SD 1.3) and 4.0 (SD 1.3, range: 0-8), respectively. There was a significant improvement in adherence across the 3-month study for both groups (mean

change 1.4, SD 1.6; *P*<.001). At 3 months, 63.2% of the intervention group compared with 55.6% of the AC group reported being adherent to their medications (MMAS-8 score≥6); however, there were no between-group differences ($F_{1,36}$=0.5; *P*=.50; Table 3).

Table 3. Change in medication adherence blood pressure and hemoglobin A1c between baseline and 3 months by study group.

Outcome	Control participants (n=21)		Intervention participants (n=21)		*F* test[a] (*df*)	*P* value
	Baseline	3 months	Baseline	3 months		
Medication adherence, mean (SD); range: 0-8[b,c]	4.0 (1.3)	5.5 (2.1)	4.4 (1.3)	5.6 (2.0)	0.5 (1, 4)	.50
Change	N/A[d]	1.5	N/A	1.2	N/A	N/A
Systolic BP[e], mean (SD) mm Hg	137.4 (17.8)	135.1 (19.5)	139.9 (18.3)	130.9 (17.4)	3.1 (1, 26)	.10
Change	N/A	−2.3	N/A	−9.0	N/A	N/A
Diastolic BP, mean (SD) mm Hg	88.5 (10.9)	87.4 (10.3)	84.1 (14.1)	80.2 (16.0)	2.9 (1, 27)	.10
Change	−1.1	N/A	−3.8	N/A	N/A	N/A
Hemoglobin A_{1c}, n (%)	7.3 (2.8)	7.8 (2.5)	8.5 (3.0)	8.2 (2.7)	1.1 (1, 30)	.30
Change	N/A	+0.5	N/A	−0.3	N/A	N/A

[a]*F* statistic results of the analysis of covariance.

[b]Higher scores indicate better adherence.

[c]The eight-item Morisky Medication Adherence Scale (MMAS-8) scoring and coding presented in the study was done using the electronic Morisky Widget MMAS-8 software. The use of the Morisky Widget MMAS-8 software, copyright registration number TX 8-816-517, is protected by US copyright laws. Permission for use of the Morisky Widget MMAS-8 software is required and was obtained for this research. A license agreement is available from MMAS Research LLC 14725 NE 20th St Bellevue, WA 98007, United States; strubow@morisky.org.

[d]N/A: not applicable.

[e]BP: blood pressure.

Effect of the Intervention on BP and HbA_{1c}

The mean baseline BP for the intervention group was 139.9 (SD 18.3)/84.1 (SD 14.4) mm Hg and that for the AC group was 137.4 (SD 17.8)/87.4 (SD 10.3) mm Hg. SBP significantly improved over time for the total sample (mean −4.8, SD 16.1 mm Hg; *P*=.04). The intervention group showed a 6.7 mm Hg greater reduction in SBP than the AC group, with no between-group difference ($F_{1,26}$=3.1; *P*=.10). The reduction in DBP across the 3 months for the total sample was nonsignificant (mean −1.97, SD 9.19 mm Hg; *P*=.20); however, the intervention group showed a 2.7 mm Hg greater reduction in DBP than the AC group.

The mean HbA_{1c} was 8.2% (SD 2.7) in the intervention group and 7.8% (SD 2.5) in the AC group. The decrease in HbA_{1c} across the 3 months for the total sample was nonsignificant (mean −0.2%, SD 0.3; *P*=.50). However, the intervention group exhibited a 0.3% reduction in HbA_{1c} over time, whereas the AC group showed a 0.5% increase.

Discussion

Principal Findings

This feasibility study evaluated the acceptability and preliminary efficacy of a theory-driven mHealth intervention that was socioculturally tailored for Black patients with uncontrolled HTN and/or T2D to address their most salient barriers to medication adherence. Exit interviews demonstrated high acceptability of the intervention with patients rating it easy to use, enjoyable, and beneficial for understanding the importance of being adherent and for learning strategies to talk to their providers about medication concerns and developing habits to routinely take their medications and make lifestyle changes. Despite high acceptability, one-third of intervention participants recommended including a health educator as an adjunct to the mHealth intervention, suggesting that some in-person contact is important and could not be replaced by the design of this intervention. Future research should test whether inclusion of an avatar-narrator who uses a storytelling approach to guide patients through the program and answer questions may be a suitable alternative to the educator.

Contrary to our hypotheses, we did not observe significant between-group differences in self-reported medication adherence and SBP reduction at 3 months. However, the results did show that patients in the intervention group exhibited greater improvements in BP and HbA_{1c} across the 3-month study than patients in the AC group. Future research should replicate this study with a larger sample size for a longer duration to determine whether these effects are of clinical significance and can be sustained over time.

Our findings are similar to those of previous studies, which found that tablet-based interventions may be an acceptable approach for addressing medication nonadherence in patients with cardiovascular diseases and T2D [48-50]. However, many

of the interventions were medication management systems that provided medication reminders, similar to text messaging programs, and did not address motivational or behavioral barriers to adherence [51,52]. One exception was My Interventional Drug-Eluting Stent Education App (MyIDEA), which was a tablet-delivered intervention designed to improve antithrombotic medication adherence among 24 patients who had a percutaneous coronary intervention [49]. The MyIDEA program combined tailored information about patients' symptoms with patient vignettes about the importance of medication adherence. Usage data, measured as the time using the MyIDEA program, suggested that the use of patient vignettes was an acceptable intervention approach for this study population. Similar to this study, the intervention group also demonstrated a greater, albeit nonsignificant, increase in medication adherence than the control group.

There are several strengths to this study. First, we included the target population in the design of the intervention to ensure that the end product incorporated the needs, skills, and preferences of the users. Second, the intervention moved beyond relying on a *single bullet approach* to improve medication adherence by using individualized profiles that identified patients' most salient adherence barriers and subsequently matched the appropriate mix of strategies to address those needs. Finally, we limited our population to nonadherent patients, thereby targeting high-risk patients who are more likely to be high users of the health care system because of uncontrolled T2D or HTN and its related complications.

Limitations

Despite these strengths, there are several reasons for the null findings of our study. These may include the small sample size and short time frame. Moreover, the use of an AC condition may have served as an intervention itself, thereby diminishing our ability to find between-group differences. We may have also failed to intervene on other important barriers to adherence that were not captured by the IMB survey or identified through feedback from participants during the formative phase of the study. Although this study comprised patients with HTN or T2D, small sample sizes in each disease state (6 patients only had a diagnosis of HTN and 5 only had a diagnosis of T2D) prohibited testing the effectiveness of the intervention in either of these subgroups. However, in the exit interviews, several participants spoke about the relative importance of the 2 diseases, often regarding T2D as more dangerous and thus considered it more important to get under control than HTN. A similar finding was documented in a qualitative study of racially diverse patients with comorbid T2D and HTN [53]. Future research should explore whether additional intervention strategies are needed to address patients' perceptions about the importance of BP control when also diagnosed as having T2D.

The intervention was also delivered only once; thus, we do not know if a higher dose would have been acceptable or led to improved outcomes. Future research should examine whether implementing the intervention in the clinic waiting room as part of regular care (eg, every 3 months for patients with uncontrolled disease) would help to prepare patients to discuss challenges with medication adherence with their provider and lead to improvements in patient activation, medication adherence, and disease control [54,55]. Finally, it is possible that participants exhibited a recall bias when rating the acceptability of the intervention because these questions were asked 3 months after the completion of the intervention. The methodological limitations of our study are similar to those documented above and in systematic reviews of mHealth interventions targeting medication adherence and reinforce the call for more methodologically rigorous studies of mHealth interventions that include larger sample sizes, are of a longer duration, and use more robust measures of medication adherence to determine the sustainable impact of these approaches on health behavior change [56]. Medication adherence was also assessed by self-report, which may have resulted in an overestimation of adherence levels. Future studies should use a more objective measure of adherence to confirm our findings. Finally, although medication adherence is a primary contributor to BP and glycemic control, other factors such as changes in lifestyle behaviors may explain our reductions in BP and HbA_{1c} but were not measured in this study.

In conclusion, this feasibility study demonstrates the acceptability of a tailored mHealth adherence intervention for Black patients with uncontrolled HTN and/or T2D. Modifications to the intervention that enhance the technical functions and streamline the IMB questionnaire should guide future evaluation of the intervention in a larger sample.

Acknowledgments

The authors would like to thank Obinna Nwafur, Julia Bails, and Lauren Jacobowitz for their assistance with this project. The MMAS-8 scoring and coding presented in this study were performed using the electronic Morisky Widget MMAS-8 software. The use of the Morisky Widget MMAS-8 software, copyright registration number TX 8-816-517, is protected by US copyright laws. Permission for the use of the Morisky Widget MMAS-8 software is required and was obtained for this research. A license agreement is available from MMAS Research LLC 14725 NE 20th St Bellevue, WA 98007, United States; strubow@morisky.org. This work was supported by a grant from Merck & Co, Inc, and NIH CTSA at NYU UL1TR001445 (principal investigator: AS).

Conflicts of Interest

AS, ML, and MB have no competing interests or financial disclosures to declare. KS and WW were paid as consultants to develop the mHealth intervention for this project.

Multimedia Appendix 1
Frequency of top two adherence barriers identified by the information-motivation-behavioral skills model of adherence survey among intervention participants.
[DOCX File , 13 KB - mhealth_v8i9e17135_app1.docx]

References

1. Fox CS, Golden SH, Anderson C, Bray GA, Burke LE, de Boer IH, American Heart Association Diabetes Committee of the Council on Lifestyle and Cardiometabolic Health, Council on Clinical Cardiology, Council on CardiovascularStroke Nursing, Council on Cardiovascular Surgery and Anesthesia, Council on Quality of Care and Outcomes Research, American Diabetes Association. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: a scientific statement from the American heart association and the American diabetes association. Diabetes Care 2015 Sep;38(9):1777-1803 [FREE Full text] [doi: 10.2337/dci15-0012] [Medline: 26246459]
2. Writing Group Members, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, American Heart Association Statistics Committee, Stroke Statistics Subcommittee. Heart disease and stroke statistics-2016 update: a report from the American heart association. Circulation 2016 Jan 26;133(4):e38-360. [doi: 10.1161/CIR.0000000000000350] [Medline: 26673558]
3. Sudano JJ, Baker DW. Antihypertensive medication use in Hispanic adults: a comparison with black adults and white adults. Med Care 2001 Jun;39(6):575-587. [doi: 10.1097/00005650-200106000-00006] [Medline: 11414262]
4. Bosworth HB, Dudley T, Olsen MK, Voils CI, Powers B, Goldstein MK, et al. Racial differences in blood pressure control: potential explanatory factors. Am J Med 2006 Jan;119(1):70.e9-70.15. [doi: 10.1016/j.amjmed.2005.08.019] [Medline: 16431192]
5. Charles H, Good CB, Hanusa BH, Chang CH, Whittle J. Racial differences in adherence to cardiac medications. J Natl Med Assoc 2003 Jan;95(1):17-27. [Medline: 12656446]
6. Hyre AD, Cohen AJ, Kutner N, Alper AB, Muntner P. Prevalence and predictors of posttraumatic stress disorder among hemodialysis patients following Hurricane Katrina. Am J Kidney Dis 2007 Oct;50(4):585-593. [doi: 10.1053/j.ajkd.2007.07.013] [Medline: 17900458]
7. Hu D, Juarez DT, Yeboah M, Castillo TP. Interventions to increase medication adherence in African-American and Latino populations: a literature review. Hawaii J Med Public Health 2014 Jan;73(1):11-18 [FREE Full text] [Medline: 24470982]
8. Kountz DS. Hypertension in ethnic populations: tailoring treatments. Clin Cornerstone 2004;6(3):39-46; discussion 47. [doi: 10.1016/s1098-3597(04)80063-2] [Medline: 15707261]
9. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 1;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
10. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
11. Mayberry LS, Mulvaney SA, Johnson KB, Osborn CY. The messaging for diabetes intervention reduced barriers to medication adherence among low-income, diverse adults with type 2. J Diabetes Sci Technol 2017 Jan;11(1):92-99 [FREE Full text] [doi: 10.1177/1932296816668374] [Medline: 27595710]
12. Nelson LA, Mulvaney SA, Gebretsadik T, Ho Y, Johnson KB, Osborn CY. Disparities in the use of a mHealth medication adherence promotion intervention for low-income adults with type 2 diabetes. J Am Med Inform Assoc 2016 Jan;23(1):12-18 [FREE Full text] [doi: 10.1093/jamia/ocv082] [Medline: 26186935]
13. Buis L, Hirzel L, Dawood RM, Dawood KL, Nichols LP, Artinian NT, et al. Text messaging to improve hypertension medication adherence in African Americans from primary care and emergency department settings: results from two randomized feasibility studies. JMIR Mhealth Uhealth 2017 Feb 1;5(2):e9 [FREE Full text] [doi: 10.2196/mhealth.6630] [Medline: 28148474]
14. Dick JJ, Nundy S, Solomon MC, Bishop KN, Chin MH, Peek ME. Feasibility and usability of a text message-based program for diabetes self-management in an urban African-American population. J Diabetes Sci Technol 2011 Sep 1;5(5):1246-1254 [FREE Full text] [doi: 10.1177/193229681100500534] [Medline: 22027326]
15. Davidson T, 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]
16. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
17. Sabin LL, Larson Williams A, Le BN, Herman AR, Viet Nguyen H, Albanese RR, et al. Benefits and limitations of text messages to stimulate higher learning among community providers: participants' views of an mhealth intervention to support continuing medical education in Vietnam. Glob Health Sci Pract 2017 Jun 27;5(2):261-273 [FREE Full text] [doi: 10.9745/GHSP-D-16-00348] [Medline: 28655802]

18. Kim S, So J. How message fatigue toward health messages leads to ineffective persuasive outcomes: examining the mediating roles of reactance and inattention. J Health Commun 2018;23(1):109-116. [doi: 10.1080/10810730.2017.1414900] [Medline: 29272208]
19. Horner GN, Agboola S, Jethwani K, Tan-McGrory A, Lopez L. Designing patient-centered text messaging interventions for increasing physical activity among participants with type 2 diabetes: qualitative results from the text to move intervention. JMIR Mhealth Uhealth 2017 Apr 24;5(4):e54 [FREE Full text] [doi: 10.2196/mhealth.6666] [Medline: 28438728]
20. Vaportzis E, Clausen MG, Gow AJ. Older adults experiences of learning to use tablet computers: a mixed methods study. Front Psychol 2018;9:1631 [FREE Full text] [doi: 10.3389/fpsyg.2018.01631] [Medline: 30233467]
21. Bouchard L, Yanez B, Dahn JR, Flury SC, Perry KT, Mohr DC, et al. Brief report of a tablet-delivered psychosocial intervention for men with advanced prostate cancer: acceptability and efficacy by race. Transl Behav Med 2019 Jul 16;9(4):629-637 [FREE Full text] [doi: 10.1093/tbm/iby089] [Medline: 30285186]
22. 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]
23. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens (Greenwich) 2008 May;10(5):348-354 [FREE Full text] [doi: 10.1111/j.1751-7176.2008.07572.x] [Medline: 18453793]
24. Morisky DE, DiMatteo MR. Improving the measurement of self-reported medication nonadherence: response to authors. J Clin Epidemiol 2011 Mar;64(3):255-7; discussion 258 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.09.002] [Medline: 21144706]
25. Berlowitz DR, Pajewski NM, Kazis LE. Intensive blood-pressure treatment and patient-reported outcomes. N Engl J Med 2017 Nov 23;377(21):2096-2097. [doi: 10.1056/nejmc1712573]
26. Resnick B, Bellg AJJ, Borrelli B, Defrancesco C, Breger R, Hecht J, et al. Examples of implementation and evaluation of treatment fidelity in the BCC studies: where we are and where we need to go. Ann Behav Med 2005 Apr(29 Suppl):46-54. [doi: 10.1207/s15324796abm2902s_8] [Medline: 15921489]
27. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH behavior change consortium. Health Psychol 2004 Sep;23(5):443-451. [doi: 10.1037/0278-6133.23.5.443] [Medline: 15367063]
28. Fisher JD, Amico KR, Fisher WA, Harman JJ. The information-motivation-behavioral skills model of antiretroviral adherence and its applications. Curr HIV/AIDS Rep 2008 Nov;5(4):193-203. [doi: 10.1007/s11904-008-0028-y] [Medline: 18838059]
29. Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychol 2006 Jul;25(4):462-473. [doi: 10.1037/0278-6133.25.4.462] [Medline: 16846321]
30. Schoenthaler AM, Butler M, Chaplin W, Tobin J, Ogedegbe G. Predictors of changes in medication adherence in blacks with hypertension: moving beyond cross-sectional data. Ann Behav Med 2016 Oct;50(5):642-652 [FREE Full text] [doi: 10.1007/s12160-016-9791-y] [Medline: 26944584]
31. Ogedegbe G, Harrison M, Robbins L, Mancuso CA, Allegrante JP. Barriers and facilitators of medication adherence in hypertensive African Americans: a qualitative study. Ethn Dis 2004;14(1):3-12. [Medline: 15002917]
32. Harkness J, Villar A, Edwards B. Translation, adaptation,design. In: Survey Methods in Multinational, Multiregional, and Multicultural Contexts. Hoboken, NJ: John Wiley & Sons; 2010:117-140.
33. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification pro task force report. Value Health 2009;12(8):1075-1083 [FREE Full text] [doi: 10.1111/j.1524-4733.2009.00603.x] [Medline: 19804437]
34. Amico KR, Toro-Alfonso J, Fisher JD. An empirical test of the information, motivation and behavioral skills model of antiretroviral therapy adherence. AIDS Care 2005 Aug;17(6):661-673. [doi: 10.1080/09540120500038058] [Medline: 16036253]
35. Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, et al. Measurement of blood pressure in humans: a scientific statement from the American heart association. Hypertension 2019 May;73(5):e35-e66. [doi: 10.1161/HYP.0000000000000087] [Medline: 30827125]
36. Amico KR, Barta W, Konkle-Parker DJ, Fisher JD, Cornman DH, Shuper PA, et al. The information-motivation-behavioral skills model of ART adherence in a deep south HIV+ clinic sample. AIDS Behav 2009 Feb;13(1):66-75 [FREE Full text] [doi: 10.1007/s10461-007-9311-y] [Medline: 17876697]
37. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [doi: 10.1016/s0738-3991(98)00116-5] [Medline: 14528569]
38. Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, et al. Association of health literacy with diabetes outcomes. J Am Med Assoc 2002;288(4):475-482. [doi: 10.1001/jama.288.4.475] [Medline: 12132978]
39. Mantwill S, Monestel-Umaña S, Schulz PJ. The relationship between health literacy and health disparities: a systematic review. PLoS One 2015;10(12):e0145455 [FREE Full text] [doi: 10.1371/journal.pone.0145455] [Medline: 26698310]

40. 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]
41. Kolb S. Grounded Theory and the Constant Comparative Method: Valid Research Strategies for Educators. University of Wisconsin. 2012. URL: https://www.researchgate.net/publication/307632469_Grounded_theory_and_the_constant_comparative_method_Valid_research_strategies_for_educators [accessed 2020-09-08]
42. Strauss A, Corbin J. Grounded theory methodology. In: The SAGE Handbook of Qualitative Research. Thousand Oaks, CA: Sage Publications; 1994.
43. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
44. Schroeder K, Fahey T, Ebrahim S. Interventions for improving adherence to treatment in patients with high blood pressure in ambulatory settings. Cochrane Database Syst Rev 2004(2):CD004804. [doi: 10.1002/14651858.CD004804] [Medline: 15106262]
45. Krousel-Wood M, Hyre A, Muntner P, Morisky D. Methods to improve medication adherence in patients with hypertension: current status and future directions. Curr Opin Cardiol 2005 Jul;20(4):296-300. [doi: 10.1097/01.hco.0000166597.52335.23] [Medline: 15956826]
46. Manias E, Williams A. Medication adherence in people of culturally and linguistically diverse backgrounds: a meta-analysis. Ann Pharmacother 2010 Jun;44(6):964-982. [doi: 10.1345/aph.1M572] [Medline: 20442356]
47. Schoenthaler A, de la Calle F, Barrios-Barrios M, Garcia A, Pitaro M, Lum A, et al. A practice-based randomized controlled trial to improve medication adherence among Latinos with hypertension: study protocol for a randomized controlled trial. Trials 2015 Jul 2;16:290 [FREE Full text] [doi: 10.1186/s13063-015-0815-x] [Medline: 26134837]
48. Lee M, Dey AK. Real-Time Feedback for Improving Medication Taking. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI'14; April 26-May 1, 2014; Toronto, Canada. [doi: 10.1145/2556288.2557210]
49. 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]
50. Lynch CP, Williams JS, Ruggiero KJ, Knapp RG, Egede LE. Tablet-aided behavioral intervention effect on self-management skills (TABLETS) for diabetes. Trials 2016 Mar 22;17:157 [FREE Full text] [doi: 10.1186/s13063-016-1243-2] [Medline: 27005766]
51. Dasgupta D, Chaudhry B, Koh E, Chawla NV. A survey of tablet applications for promoting successful aging in older adults. IEEE Access 2016;4:9005-9017. [doi: 10.1109/access.2016.2632818]
52. Aldeer M, Javanmard M, Martin R. A review of medication adherence monitoring technologies. Appl Syst Innov 2018 May 6;1(2):14. [doi: 10.3390/asi1020014]
53. Stewart J, Brown K, Kendrick D, Dyas J, Nottingham Diabetes and Blood Pressure Study Group. Understanding of blood pressure by people with type 2 diabetes: a primary care focus group study. Br J Gen Pract 2005 Apr;55(513):298-304 [FREE Full text] [Medline: 15826438]
54. Grant RW, Lyles C, Uratsu CS, Vo MT, Bayliss EA, Heisler M. Visit planning using a waiting room health IT tool: the aligning patients and providers randomized controlled trial. Ann Fam Med 2019 Mar;17(2):141-149 [FREE Full text] [doi: 10.1370/afm.2352] [Medline: 30858257]
55. Sherwin HN, McKeown M, Evans MF, Bhattacharyya OK. The waiting room 'wait': from annoyance to opportunity. Can Fam Physician 2013 May;59(5):479-481 [FREE Full text] [Medline: 23673581]
56. DeKoekkoek T, Given B, Given CW, Ridenour K, Schueller M, Spoelstra SL. mHealth SMS text messaging interventions and to promote medication adherence: an integrative review. J Clin Nurs 2015 Oct;24(19-20):2722-2735. [doi: 10.1111/jocn.12918] [Medline: 26216256]

Abbreviations

AC: attention control
BP: blood pressure
DBP: diastolic blood pressure
HbA_{1c}: hemoglobin A_{1c}
HTN: hypertension
IMB: Information-Motivation-Behavioral skills model of adherence
mHealth: mobile health
MMAS: Morisky Medication Adherence Scale
MyIDEA: My Interventional Drug-Eluting Stent Education App
RA: research assistant

RCT: randomized controlled trial
s-TOFHLA: 36-item short-form Test of Functional Health Literacy in Adults
SBP: systolic blood pressure
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 20.11.19; peer-reviewed by L Nelson, J Magnani; comments to author 10.02.20; revised version received 06.07.20; accepted 11.08.20; published 23.09.20.

Please cite as:
Schoenthaler A, Leon M, Butler M, Steinhaeuser K, Wardzinski W
Development and Evaluation of a Tailored Mobile Health Intervention to Improve Medication Adherence in Black Patients With Uncontrolled Hypertension and Type 2 Diabetes: Pilot Randomized Feasibility Trial
JMIR Mhealth Uhealth 2020;8(9):e17135
URL: http://mhealth.jmir.org/2020/9/e17135/
doi:10.2196/17135
PMID:32965230

Original Paper

Feasibility of a Novel Mobile C-Reactive Protein–Testing Device Using Gold-Linked Electrochemical Immunoassay: Clinical Performance Study

Yuko Gondoh-Noda[1], MD; Mitsuhiro Kometani[2], MD, PhD; Akihiro Nomura[3,4], MD, PhD; Daisuke Aono[2], MD, PhD; Shigehiro Karashima[2], MD, PhD; Hiromi Ushijima[5], PhD; Eiichi Tamiya[5,6], PhD; Toshinori Murayama[4], MD, PhD; Takashi Yoneda[2,7,8], MD, PhD

[1]Department of Clinical Development, Kanazawa University Graduate School of Medicine, Kanazawa, Japan

[2]Division of Endocrinology and Hypertension, Department of Cardiovascular and Internal Medicine, Graduate School of Medical Sciences, Kanazawa University, Kanazawa, Japan

[3]Department of Cardiovascular Medicine, Kanazawa University Graduate School of Medical Sciences, Innovative Clinical Research Center, Kanazawa University (iCREK), Kanazawa, Japan

[4]Innovative Clinical Research Center, Kanazawa University (iCREK), Kanazawa, Japan

[5]BioDevice Technology Ltd, Nomi, Japan

[6]Department of Applied Physics, Graduate School of Engineering, Osaka University, Nomi, Japan

[7]Department of Health Promotion and Medicine of the Future, Graduate School of Medical Sciences, Kanazawa University, Kanazawa, Japan

[8]Program Management Office for Medical Innovation, Graduate School of Medical Sciences, Kanazawa University, Kanazawa, Japan

Corresponding Author:
Takashi Yoneda, MD, PhD
Division of Endocrinology and Hypertension, Department of Cardiovascular and Internal Medicine
Graduate School of Medical Sciences
Kanazawa University
13-1 Takara-machi
Kanazawa, 920-8641
Japan
Phone: 81 76 265 2778
Email: endocrin@med.kanazawa-u.ac.jp

Abstract

Background: Home-based care is one of the most promising solutions to provide sufficient medical care for several older patients in Japan. However, because of insufficient diagnostic devices, it is sometimes difficult to detect early signs of the occurrence or worsening of diseases, such as infections under home-based care settings. C-reactive protein (CRP) is highly sensitive to diagnosing infections, and its elevation can help diagnose acute infection in older patients. Therefore, a CRP-measuring device that can be used in such a specific occasion is needed for home-based care. However, aspects such as its size, weight, and procedure are still challenging with respect to the practical use of mobile devices that quantitatively measure CRP levels easily and quickly under home-based care settings.

Objective: We developed a new mobile, rapid CRP measurement device using a gold-linked electrochemical immunoassay (GLEIA) system. The aim of this study was to evaluate the feasibility of this mobile CRP-testing device.

Methods: First, we assessed the performance of bare GLEIA-based electrode chips as the foundation of the device. After embedding the bare GLEIA-based electrode chips in a special plastic case and developing the mobile CRP-testing device, we further tested the device prototype using clinical blood samples. Finally, we evaluated the intra-assay variability for precision in the same condition and inter-assay variability for reproducibility in different conditions.

Results: Blood samples for analysis were obtained by direct vein puncture from outpatients (N=85; females: 57/85; males: 28/85; age: 19-88 years) at Kanazawa University Hospital in Japan. For performance evaluation of bare GLEIA-based electrode chips, we used 85 clinical blood samples. There was a significant positive correlation between the electrode-predicted CRP levels and the reference CRP concentrations (R^2=0.947; P<.001). The assembled device was mobile (size 45×90×2.4 mm; weight 10 g) and disposable. The minimum volume of the sample needed for measuring CRP was 1.4 µL. The estimated preanalytical time

was approximately 7 minutes and 40 seconds, and analysis time was approximately 1 minute and 10 seconds. Subsequently, for performance evaluation of the mobile CRP-testing device using GLEIA-based electrode chips, we used 26 clinical blood samples and found a significant positive correlation between the mobile device-predicted CRP levels and the reference CRP concentrations (R^2=0.866, P<.001). The intra-assay variabilities were 34.2%, 40.8%, and 24.5% for low, medium, and high CRP concentrations, respectively. The inter-assay variabilities were 46.5%, 38.3%, and 64.1% for low, medium, and high CRP concentrations, respectively.

Conclusions: Our findings suggest that this new mobile CRP-testing device might be suitable for use in home-based care settings.

(JMIR Mhealth Uhealth 2020;8(9):e18782) doi:10.2196/18782

KEYWORDS

gold-linked electrochemical immunoassay (GLEIA); home-based care; mobile CRP testing device; mHealth; diagnostic; infection; assay; CRP; c-reactive protein; immunoassay

Introduction

Home-based care is one of the most promising solutions to provide sufficient medical care for the large number of older patients in Japan, because the ageing rate in Japan is expected to reach as high as 26.2% in 2020 [1]. Home-based care mainly provides general medical care for older patients who face difficulties in seeing doctors at hospitals or those who wish to live in their homes rather than at medical facilities. Home-based care also meets the governmental demand in Japan to alleviate hospitalization burden and to decrease total medical costs [2]. However, because of insufficient diagnostic devices, it is sometimes difficult to detect early signs of disease occurrence or worsening such as by infections in these settings [2-4]. Diagnostic delay worsens the prognosis of infectious diseases [5]. Thus, a better tool for early diagnosis of infection is needed in home-based care settings.

Nowadays, blood tests to check systemic inflammation are being used in Japanese hospitals, as they are easy and useful for early diagnosis of infectious diseases [6-8]. C-reactive protein (CRP), a plasma protein and a major component of inflammatory reactions, is widely used as an objective indicator of systemic inflammation [9-11]. CRP is highly sensitive for diagnosing infections [12], and its elevation can help diagnose acute infection in older patients [13]. It is also known that CRP is a useful tool for correlation with longer overall survival in early-stage malignancies [14]. Although at least 8 different semi-quantitative strips and quantitative point-of-care tests for CRP level have already been reported, each of these methods has both practical merits and demerits [15]. One of them is too big and heavy to carry, and the others need long times to obtain the result or have limitations related to their material.

The gold-linked electrochemical immunoassay (GLEIA) system is a highly sensitive electrochemistry assay that uses gold nanoparticle–labeled antibodies [16], and provides advantages in the aspects of miniaturization and time saving. Therefore, we developed a new mobile and rapid CRP measurement device using a GLEIA system for quantitative, easy, and immediate measurement of serum CRP levels from patients' blood samples. In this study, we evaluated the feasibility of a new mobile, rapid CRP measurement device using the GLEIA system.

Methods

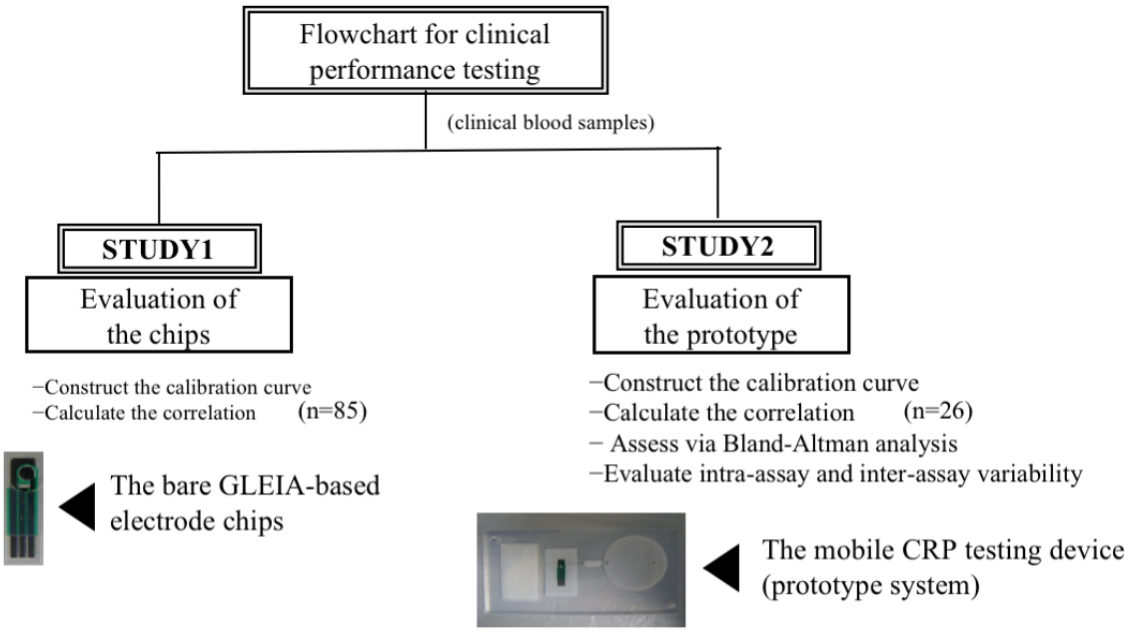
Overview

Figure 1 shows the flowchart of clinical performance tests for the mobile CRP-testing device.

In study 1, we first assessed the electrolytic currents associated with standardized CRP concentrations (0, 0.25, 1, 4, 8, and 16 mg/dL) using bare GLEIA-based electrode chips and constructed a "current-to-CRP" curve for calibration. Next, we established a prediction formula using the calibration curve to determine the serum CRP concentration from the electrolytic current. We then recruited 85 participants to validate the bare GLEIA-based electrode chips using clinical specimens. The participants underwent a blood test, and their CRP level was checked in advance at Kanazawa University Hospital for suspected infectious diseases or to evaluate the condition of a disease. We calculated the correlation between chip-based measurements and laboratory measurements. After developing the mobile CRP-testing device (prototype system) embedded with the GLEIA-based electrode chips, we repeated the verification procedure using that prototype system (study 2). In study 2, we assessed the electrolytic currents associated with standardized CRP concentrations (0, 0.25, 1, 4, 8,16, and 32 mg/dL) using the prototype device and constructed a "current-to-CRP" curve for calibration. We then validated the prototype system using 25 clinical samples as in study 1. We also performed Bland-Altman analysis and intra- and inter-assay variability testing of the mobile CRP-testing device for practical use.

This study was approved by the Ethics Committees of Kanazawa University (no. 2017078). The written informed consent was waived because we used existing samples with no additional invasion to the patients. We used an opt-out approach to protect the patients' rights to reject the participate in this study. This was done by posting the study description document on the website of Kanazawa University Hospital, which is open to every patient. All methods were performed in accordance with the approved guidelines and regulations.

Figure 1. Flowchart of the clinical performance tests. CRP: c-reactive protein; GLEIA: gold-linked electrochemical immunosorbent assay.



Blood Samples

Blood samples were obtained by direct vein puncture from outpatients (N=85; females: 57/85; male: 28/85; age: 19-88 years) at Kanazawa University Hospital in Japan. Their CRP concentrations ranged from 0.0 to 15.5 mg/dL. These blood samples were collected in blood collection tubes with sodium citrate and were centrifuged at 3000*g* for 10 minutes within 1 day after collection. We used 100 µL of serum from these blood samples for the GLEIA-based measurement of CRP levels. Serum was transferred to microtubes and frozen at –80°C, and it took approximately 1 week before the serum was frozen. Frozen serum samples packed in a dedicated container were properly transported with ice packs to the laboratory of BioDevice Technology Inc by the coinvestigators. Serum samples were thawed and analyzed using the bare GLEIA-based electrode chips and the prototype system.

The GLEIA System

In the GLEIA method, immobilized primary antibodies on electrodes and gold nanoparticle–linked secondary antibodies form a sandwich structure with the antigen. Anti–C-reactive protein antibody (RRID: AB_2085618, Abcam) and Gold nanoparticles of 60 nm (Gold nanoparticles 60 nm, BBI Solutions) were used in this study. After the reaction, the free gold-linked antibodies were removed by washing with an acidic solution to fix the gold nanoparticles on the electrode surface and to oxidize the gold nanoparticles. After measuring the reduction current of oxidized gold nanoparticles using differential pulse voltammetry (DPV), we quantified the CRP levels using a "current-to-CRP" calibration curve, which was generated using quality control serum samples. We used the equation of the calibration curve to determine the CRP concentrations. The results of the test were not blinded to the operators performing the GLEIA measurements.

All 85 samples were analyzed using GLEIA electrode chips to determine the correlation between CRP levels (measured using the bare GLEIA-based electrode chips) and the laboratory-measured CRP concentrations as reference. Initially, we measured the reduction current; if the current exceeded the maximum value, it was rechecked up to 3 times. Data were rechecked in 7 cases. Some samples were rechecked at least once. Following this rechecking process, 4 samples were excluded from further analyses because no data were obtained.

After developing the new GLEIA measurement system, we used 26 samples for further analysis using the prototype. If the current exceeded the maximum value, the data were rechecked; one sample showed no data.

Performance Evaluation of the Bare GLEIA-Based Electrode Chips

Initially, serum was added to a microtube containing gold-linked secondary antibodies; these serum samples were then mixed with the diluting solution and diluted 1000-fold. The samples were then placed on the GLEIA electrode chips immobilized primary antibodies, and after formation of the sandwich structure resulting from the antigen-antibody reaction, excess antibodies were washed out by rinsing with the washing solution in a beaker. Finally, we connected the analyzer which uses DPV system to the electrode chips in order to determine the results and calculated the CRP levels using an Internet of Things (IoT) device. This IoT device is a computer terminal using proprietary software for the analyzer (BDTminiSTAT100) and the environment is Windows 7 to 10.

Performance Evaluation of the Mobile CRP-Testing Device

The mobile device size was 45×90×2.4 mm, and its weight was 10 g. The analyzer (BDTminiSTAT100) was 50×70×25 mm in size, and it weighed 65 g. The adaptation equipment size was

70×120×15 mm, and it weighed 130 g. This adaptation equipment is necessary to connect the electrode to the analyzer in the correct manner. Initially, serum was added to wells coated with the gold-linked antigen, and then serum samples were mixed with the diluting solution and diluted 1000-fold. These samples were then placed on the GLEIA electrode, and after formation of the sandwich structure resulting from the antigen-antibody reaction, excess antibodies were washed out by attaching a liquid tank to the prototype system. Finally, we connected the analyzer to determine the results and calculated the CRP levels using an IoT device for Windows.

Laboratory Measurements

Results from laboratory measurements were determined before the results from GLEIA measurements. We used Quoligent CRP reagent (Sekisui Medical Co, Ltd) and a Hitachi LABOSPECT-L instrument for laboratory measurements and then visualized the results using electronic medical records within a few hours. The analytical range of the measurements was between 0.02 and 42 mg/dL, and the intra-assay coefficient of variation was less than or equal to 5%.

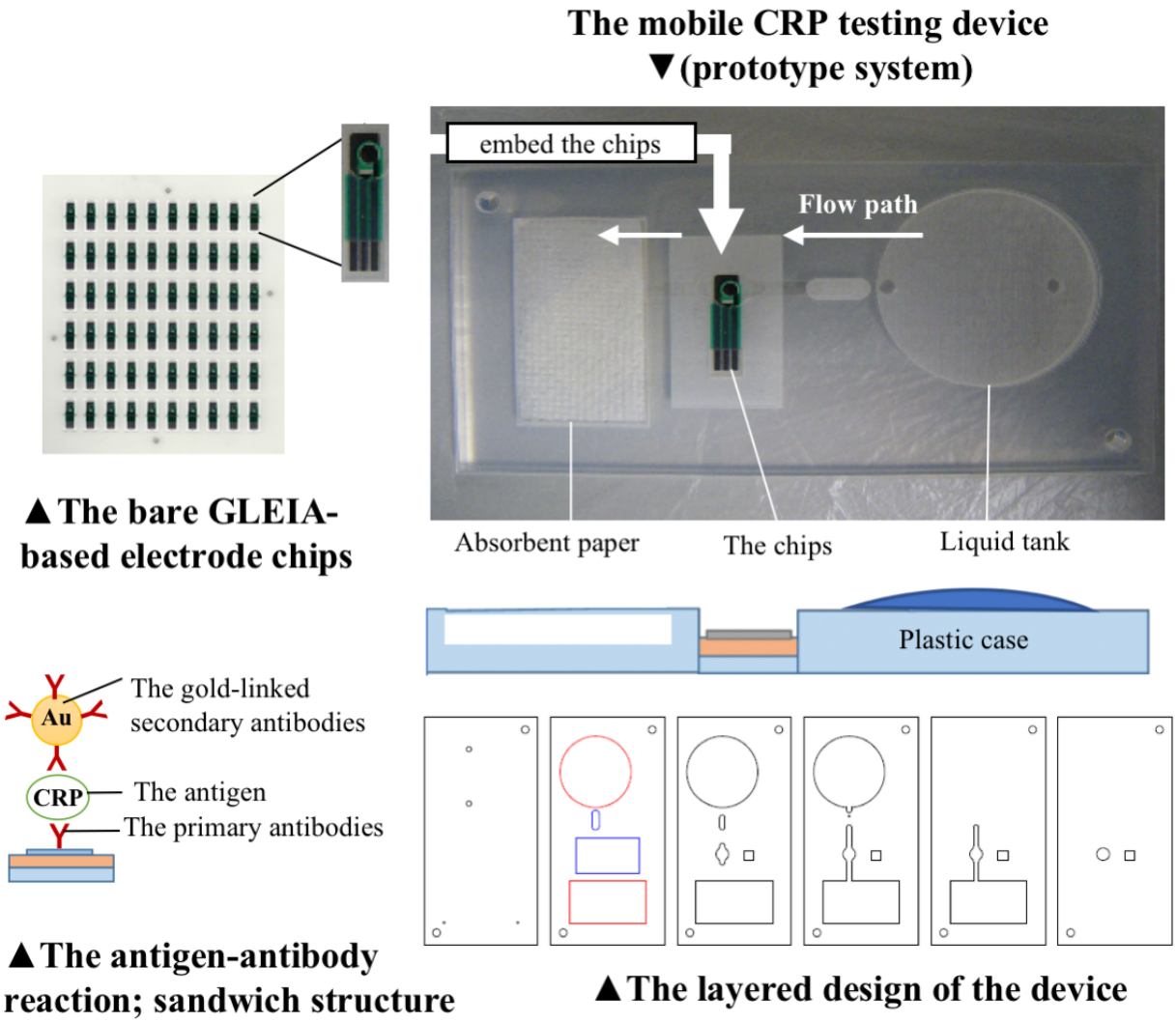
Statistical Analysis

Data were expressed as means (SD). Statistical analyses were performed using Excel 2016 (Microsoft) with an add-in software Statcel4 (OMS). Spearman rank correlation coefficient was used to analyze correlations between GLEIA measurements and laboratory measurements, and *P* values less than 0.05 were considered statistically significant. Bland-Altman plots were drawn using R version 3.4.3.

Results

The production process from the chips to the prototype and the diagram depicting the reaction occurring on the chip are shown in Figure 2. The mobile CRP-testing device (prototype system) was embedded with the GLEIA-based electrode chips and the liquid tank. The diagram depicting the reaction occurring on the chip is shown at the lower left. The cross-section of plastic case is shown at the lower right.

Figure 2. Production process from the chips to the prototype and the schematic of GLEIA measurement. CRP: c-reactive protein; GLEIA: gold-linked electrochemical immunosorbent assay.

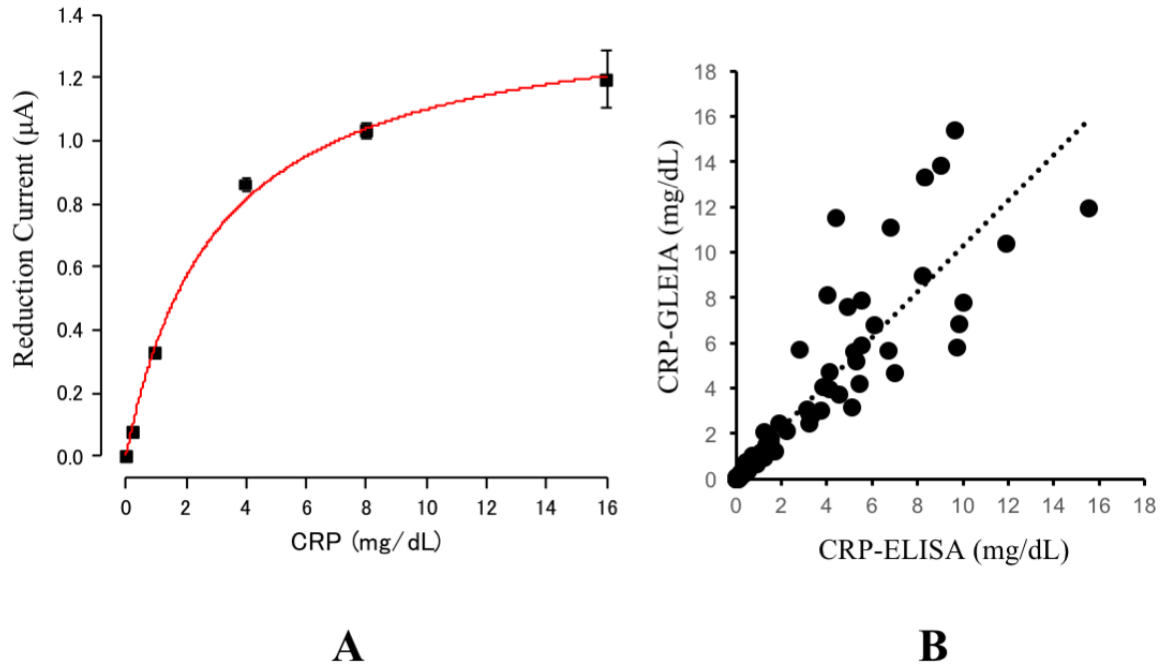


Study 1: Performance of the Bare GLEIA-Based Electrode Chips

First, we tested the performance of bare GLEIA-based electrode chips. We constructed an electrolytic reduction current-to-CRP calibration curve based on the electrolytic current produced by the chips and each standardized CRP concentration using the Michaelis-Menten model [17]. Figure 3 A represents the calibration curve between reduction current (μA) and CRP (mg/dL) level measured using GLEIA electrode chips. The formula for the Michaelis-Menten model between the current (μA) and CRP (mg/dL) was as follows: $y = V_{max}*x / (K_m + x)$, $V_{max} = 1.43697$, $K_m = 30.56623$.

For performance evaluation, we measured the reduction current using 85 clinical blood samples. If the current exceeded the maximum value, it was rechecked at least once because it did not fit within the calibration curve. Following this rechecking process, 4 samples were excluded from further analyses because no data were obtained. Then, we tested the correlation between the electrode-predicted CRP levels (based on the prediction formula) and the laboratory-measured CRP concentrations (used as reference). Figure 3 B presents the correlation between CRP levels measured using the GLEIA electrode chips and the laboratory-measured CRP concentrations (reference). There was a significant positive correlation (R=0.972962; n=81; y=1.0049 + 0.2305x; *P*<.001) between the CRP levels detected by these 2 methods. There was a significant positive correlation between the electrode-predicted CRP levels and the reference CRP concentrations (R^2=0.947; *P*<.001).

Figure 3. The current-to-CRP calibration curve (A) and the correlation between the chips-measured CRP levels and the laboratory-measured CRP levels (B). CRP: c-reactive protein; ELISA: enzyme-linked immunosorbent assay.



Study 2: Performance of the Mobile CRP-Testing Device Using the GLEIA-Based Electrode Chips

Next, we assessed the performance of the mobile CRP-testing device using the GLEIA-based electrode chips tested in study 1. The device was mobile (size 45×90×2.4 mm; weight 10 g) and disposable; and the adaptation equipment could be used several times. The minimum amount of material required for measuring CRP was 1.4 μL. The estimated preanalytical time was approximately 7 minutes and 40 seconds, and the analysis time was approximately 1 minute and 10 seconds.

We again constructed an electrolytic reduction current-to-CRP calibration curve between the electrolytic current produced by the mobile CRP-testing device and each standardized CRP concentration using the Michaelis-Menten model. Figure 4 A presents the calibration curve between the reduction current (μA) and CRP (mg/dL) using the prototype mobile CRP-testing device. The CRP level prediction formula based on the calibration curve was as follows: $y = V_{max}*x / (K_m + x)$, $V_{max} = 1.56341$, Km = 74.54069.

For performance evaluation, we measured the reduction current using 26 clinical blood samples. Of these, 1 sample yielded no data and was excluded from further analyses. We then tested the correlation between electrode-predicted CRP levels using the prediction formula and the laboratory-measured CRP concentrations as a reference. Figure 4 B presents Correlation between the mobile CRP-testing device and the laboratory-measured CRP concentrations (reference). There was a significant positive correlation (R=0.930769; n=25; y=0.898x + 0.6919; *P*<.001) between the CRP levels detected by these 2 methods. As expected, we found a significant positive correlation between the mobile device-predicted CRP levels and the reference CRP concentrations (R^2=0.866, *P*<.001).

Furthermore, we performed Bland-Altman analysis for elucidating whether there were any systematic errors. The results of the agreement between the CRP level using the prototype and laboratory measurements are shown graphically in Figure 5. Differences between the CRP levels measured using the mobile CRP-testing device and the laboratory-measured CRP concentrations were calculated for each method and were plotted against the mean values of both measurements. The mean difference was 0.234. On an average, higher values (CRP > 10 mg/dL) tended to exhibit greater discrepancies between the values compared to the lower values.

Figure 4. The current-to-CRP calibration curve (A) and the correlation between the prototype-measured CRP levels and the laboratory-measured CRP levels (B). CRP: c-reactive protein; ELISA: enzyme-linked immunosorbent assay.

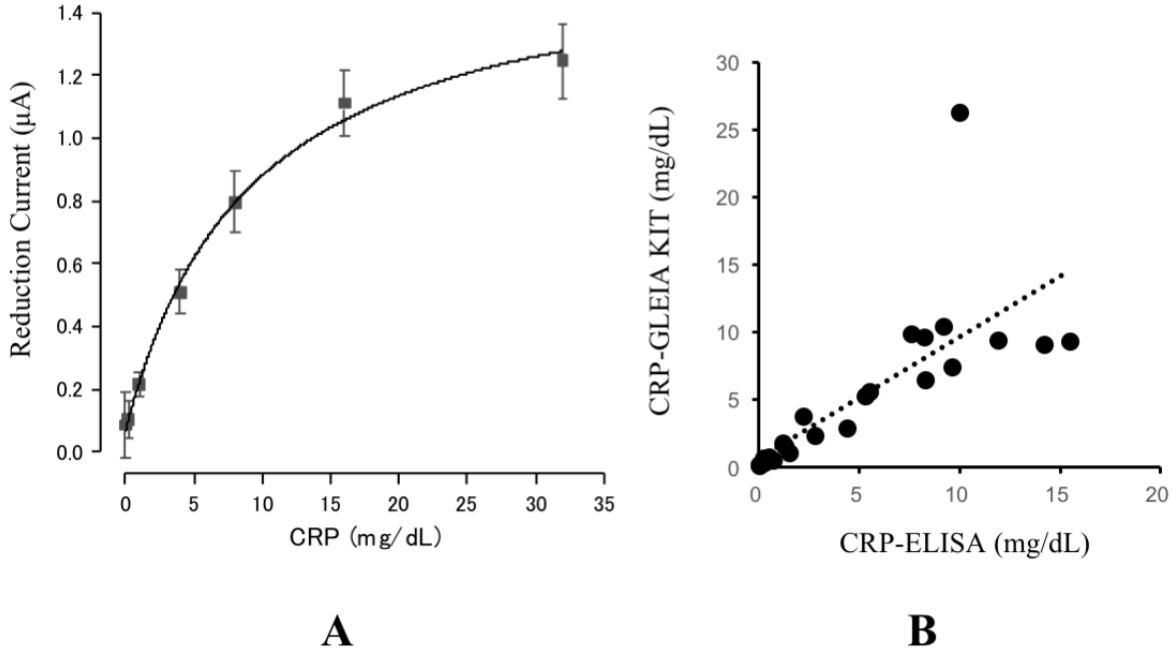
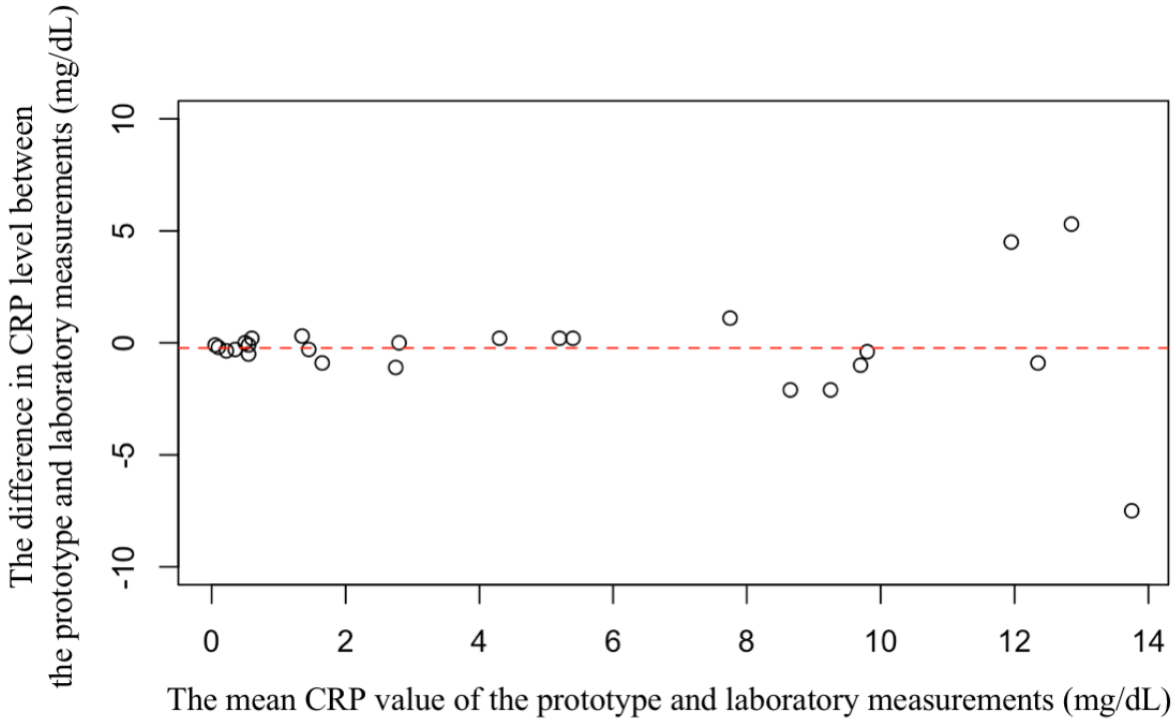


Figure 5. Bland-Altman plot between GLEIA measurement and the laboratory-measured CRP concentrations (reference). CRP: c-reactive protein.



Intra- and Inter-Assay Variability Assessments

Finally, we evaluated the (1) intra-assay variability for precision in the same condition and (2) inter-assay variability for reproducibility in different conditions. First, intra-assay variability of the prototype system was assessed by determining the coefficient of variation using 5 replicate measurements for each of the following 3 samples (low, medium, and high CRP concentration levels) on the same day and in the same conditions. These 3 samples were selected from the rest of the above 85 specimens in terms of CRP concentration for matching approximately low (0 mg/dL), medium (5 mg/dL), and high (over 10 mg/dL) CRP concentration levels, respectively. The intra-assay variabilities were 34.2% for the low (mean 0.5 mg/dL), 40.8% for the medium (5.1 mg/dL), and 29.5% for the high CRP concentrations (14.7 mg/dL).

Second, inter-assay variability of the prototype system was assessed by determining the coefficient of variation using 15 replicate measurements for each of the above 3 samples (low, medium, and high CRP concentration levels) in different conditions. We checked the same sample 3-times on 5 days with different technicians and devices (miniSTAT/pipette). The inter-assay variabilities were 46.5% for the low (mean 0.5 mg/dL), 38.3% for the medium (5.1 mg/dL), and 64.1% for the high CRP concentrations (14.7 mg/dL).

Discussion

Principal Results

In this study, we developed and evaluated the feasibility of a new mobile rapid CRP measurement device using the GLEIA system. Comparison with conventional measurement using enzyme-linked immunosorbent assay (ELISA) has often been performed to evaluate new devices [18]. Therefore, we first assessed the performance of bare GLEIA-based electrode chips and found a significant positive correlation between the electrode-predicted CRP levels and the reference CRP concentrations. Next, we assembled and tested the mobile CRP-testing device (prototype system) embedded with the GLEIA-based electrode and found a significant positive correlation between the mobile device-predicted CRP levels and the reference CRP concentrations. We also found the limitation of increasing discrepancy at higher levels of CRP. There is a possibility to reduce the discrepancy by optimizing the dilution of the samples. Further studies are needed to use this new mobile device for quantitative measurement of CRP levels easily and quickly in home-based care settings.

This study yielded several results. The GLEIA-based electrode chip was feasible for measuring CRP levels using clinical specimens. Initially, the GLEIA-based electrode chip was mass-produced in a sheet form. Therefore, we generated calibration curves for each measurement to avoid differences between electrodes. As these electrodes are disposable, there is no need to worry about contamination; and their flat shape makes it easy to modify and fix them to a device. This GLEIA system measures the reduction current of oxidized gold nanoparticles, which forms a sandwich structure with the antigen, through an antigen-antibody reaction; here, we quantified CRP levels using a “current-to-CRP” calibration curve. Thus, it takes a short time to measure the sample and the analyzer needs not to be large. We guaranteed the performance of the bare GLEIA-based electrode chips, which serve as a foundation for the mobile device. However, for the measurement procedure, we had to prepare a washing solution to remove the excess labeled antibody after the antigen-antibody reaction. We performed a rinsing step using this solution when the bare GLEIA-based electrode chips alone were used. This process is difficult in home-based care settings. Hence, we combined the GLEIA original electrode chips and a liquid tank to simplify the procedure. As a result, the rinsing step is completed within the device; the solution is pushed out through the flow path to wash out the electrodes and is imbibed by the equipped absorbent paper. This avoids the need for additional equipment and reduces the risk of contamination. The combined device was compact and lightweight, and similar in size to a name card. To our knowledge, such a lightweight device has not been reported yet [14]. Moreover, this system (patent pending) reduces the necessity for complicated procedures and avoids the risk of mistakes during artificial manipulation. However, quality assurance and training protocols need to be established to ensure maximal benefits for patient care and efficiency [19].

Further, our novel mobile CRP detector using the GLEIA-based electrode chip system was also feasible for measuring serum CRP levels. After the foundation of the device was prepared using plastic, the liquid tank was filled with the washing solution, and the device was ready to use, we confirmed that the CRP levels measured using new mobile rapid CRP-testing device also showed a significant positive correlation with the reference CRP concentrations. We then connected this device with an analyzer connected to a PC for visualizing the results on a PC screen. Thus, this device was also applicable for use with the IoT approach [20]. Alternatively, the results could be observed on a smart device if a Bluetooth-compatible analyzer were used simultaneously. This approach may be useful for data storage, management, and telemedicine, allowing us to share data with a physician at a distant location.

Limitations

This study has some limitations. First, we investigated GLEIA using clinical specimens that were already subjected to centrifugal separation in a laboratory. In home-based care settings, we hope to be able to perform the test with a simple fingertip prick, rather than with serum samples obtained after centrifugation. Second, detection of high CRP levels by the device resulted in more dramatic differences between 2 measurements than in the case of low CRP levels, and some cases needed retesting or were unmeasurable. This may be due to the use of a calibration curve that plateaus at high CRP levels. All 4 samples, which had no data in this study, showed a reduction current over 1.2 μA; and the reference CRP levels using ELISA for these 4 samples were 9.2, 13.3, 7.6, and 14.2 mg/dL, respectively. We calculated the correlation without these unmeasurable samples. Therefore, we have to mention that this correlation does not focus on high ranges of CRP level. We thus have to avoid situations wherein results showing high CRP levels may not be accurate and could therefore result in incorrect clinical decisions. Given the nature of this cause, further improvement of this approach could be theoretically achieved

by optimizing the sample dilution. Using the current methods, the serum samples were mixed with the diluting solution and diluted 1000-fold. However, if they are diluted up to 10,000-fold, the calibration curve cannot form a plateau even at a high CRP range, thus theoretically reducing errors. There is no data from a similar mobile CRP-testing device for comparison; therefore, we need to improve the results of intra- and inter-assay variabilities using the same solution. However, in home-based care settings, cases with CRP levels over 4.35 mg/dL and those with suspected pneumonia [21] (and exceeding this level) need prompt medical attention. Thus, there is a possibility to contrive the unmeasurable limitation as an indication of "extremely high CRP level."

In this study, we established a new mobile rapid CRP-testing device using a GLEIA system. There was a significant correlation between the CRP levels measured using the new mobile CRP-testing device and the laboratory-measured CRP concentrations. Because of its portability, we consider that the device might be suitable for use in home-based care settings.

Comparison With Prior Work

To date, at least 9 different types of semiquantitative strips and quantitative point-of-care testing devices have been developed for measuring CRP levels; these approaches include immunochromatographic assays, immunoturbidimetric assays, solid-phase immunochemical assays, and vertical flow assays with 3D paper-based microfluidics [22]. In some products, the analyzer weights ranged from 1.7 to 35 kg. These products were not suitable to be carried and used in home-based care settings, and other assays had limitations in their clinical use.

Conclusions

The GLEIA-based mobile device developed in this study allows quantitative measurement of serum CRP levels from patients' blood samples easily and immediately. Our findings indicate that this new mobile CRP-testing device could be suitable for use in home-based care settings. Further research is needed to apply this device to more user-friendly device or multi-item measurement.

Acknowledgments

We thank Rie Oka and Yoshihiro Watanabe for their advice regarding this research and Asahi Rubber Inc. (Saitama, Japan) for making this prototype system. We also thank the patients for providing their clinical information and samples and Yukiko Takamiya and Chiyoko Takata for their secretarial assistance. This study was supported by a research grant from Kanazawa University Hospital. This work was supported by a Clinical Research Grant from Kanazawa University Hospital, Ishikawa, Japan.

Authors' Contributions

YGN, MK, AN, SK, DA, and TY designed and supervised the experiments. MK and AN performed data analysis. HU and ET developed the device. YGN and HU collected and evaluated the data. TM and supervised the experiments. YGN, MK, and AN wrote the paper. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Anderson GF, Hussey PS. Population aging: a comparison among industrialized countries. Health Aff (Millwood) 2000;19(3):191-203. [doi: 10.1377/hlthaff.19.3.191] [Medline: 10812799]
2. Arai H, Ouchi Y, Yokode M, Ito H, Uematsu H, Eto F, Members of Subcommittee for Aging. Toward the realization of a better aged society: messages from gerontology and geriatrics. Geriatr Gerontol Int 2012 Jan;12(1):16-22. [doi: 10.1111/j.1447-0594.2011.00776.x] [Medline: 22188494]
3. Gavazzi G, Krause K. Ageing and infection. Lancet Infect Dis 2002 Nov;2(11):659-666. [doi: 10.1016/s1473-3099(02)00437-1] [Medline: 12409046]
4. Yoshikawa TT. Epidemiology and unique aspects of aging and infectious diseases. Clin Infect Dis 2000 Jun;30(6):931-933. [doi: 10.1086/313792] [Medline: 10880303]
5. Smith IM. Infections in the elderly. Hosp Pract (Off Ed) 1982 Jul;17(7):69-77, 81. [doi: 10.1080/21548331.1982.11698074] [Medline: 6809573]
6. Ishida T, Tachibana H, Ito A, Yoshioka H, Arita M, Hashimoto T. Clinical characteristics of nursing and healthcare-associated pneumonia: a Japanese variant of healthcare-associated pneumonia. Intern Med 2012;51(18):2537-2544 [FREE Full text] [doi: 10.2169/internalmedicine.51.7987] [Medline: 22989823]
7. Costello-White R, Ryff CD, Coe CL. Aging and low-grade inflammation reduce renal function in middle-aged and older adults in Japan and the USA. Age (Dordr) 2015 Aug;37(4):9808 [FREE Full text] [doi: 10.1007/s11357-015-9808-7] [Medline: 26187318]
8. Miura F, Okamoto K, Takada T, Strasberg SM, Asbun HJ, Pitt HA, et al. Tokyo Guidelines 2018: initial management of acute biliary infection and flowchart for acute cholangitis. J Hepatobiliary Pancreat Sci 2018 Jan;25(1):31-40. [doi: 10.1002/jhbp.509] [Medline: 28941329]

9. Wyczalkowska-Tomasik A, Czarkowska-Paczek B, Zielenkiewicz M, Paczek L. Inflammatory Markers Change with Age, but do not Fall Beyond Reported Normal Ranges. Arch Immunol Ther Exp (Warsz) 2016 Jul;64(3):249-254 [FREE Full text] [doi: 10.1007/s00005-015-0357-7] [Medline: 26283530]
10. Ansar W, Ghosh S. Inflammation and inflammatory diseases, markers, and mediators: Role of CRP in some inflammatory diseases. In: Biology of C Reactive Protein in Health and Disease. New Delhi: Springer; Mar 24, 2016:67-107.
11. Ansar W, Ghosh S. C-reactive protein and the biology of disease. Immunol Res 2013 May;56(1):131-142. [doi: 10.1007/s12026-013-8384-0] [Medline: 23371836]
12. Hu L, Shi Q, Shi M, Liu R, Wang C. Diagnostic Value of PCT and CRP for Detecting Serious Bacterial Infections in Patients With Fever of Unknown Origin: A Systematic Review and Meta-analysis. Appl Immunohistochem Mol Morphol 2017 Sep;25(8):e61-e69. [doi: 10.1097/PAI.0000000000000552] [Medline: 28885233]
13. Ticinesi A, Lauretani F, Nouvenne A, Porro E, Fanelli G, Maggio M, et al. C-reactive protein (CRP) measurement in geriatric patients hospitalized for acute infection. Eur J Intern Med 2017 Jan;37:7-12. [doi: 10.1016/j.ejim.2016.08.026] [Medline: 27594414]
14. Deme D, Telekes A. Prognostic importance of plasma C-reactive protein (CRP) in oncology. Orv Hetil 2017 Mar;158(7):243-256. [doi: 10.1556/650.2017.30646] [Medline: 28462626]
15. Brouwer N, van Pelt J. Validation and evaluation of eight commercially available point of care CRP methods. Clin Chim Acta 2015 Jan 15;439:195-201. [doi: 10.1016/j.cca.2014.10.028] [Medline: 25445415]
16. Xuan Viet N, Chikae M, Ukita Y, Maehashi K, Matsumoto K, Tamiya E, et al. Gold-linked electrochemical immunoassay on single-walled carbon nanotube for highly sensitive detection of human chorionic gonadotropin hormone. Biosens Bioelectron 2013 Apr 15;42:592-597. [doi: 10.1016/j.bios.2012.11.017] [Medline: 23261694]
17. Kolomeisky AB. Michaelis-Menten relations for complex enzymatic networks. J Chem Phys 2011 Apr 21;134(15):155101 [FREE Full text] [doi: 10.1063/1.3580564] [Medline: 21513417]
18. Fukunaga S, Kuwaki K, Mitsuyama K, Takedatsu H, Yoshioka S, Yamasaki H, et al. Detection of calprotectin in inflammatory bowel disease: Fecal and serum levels and immunohistochemical localization. Int J Mol Med 2018 Jan;41(1):107-118 [FREE Full text] [doi: 10.3892/ijmm.2017.3244] [Medline: 29115397]
19. Larsson A, Greig-Pylypczuk R, Huisman A. The state of point-of-care testing: a European perspective. Ups J Med Sci 2015 Mar;120(1):1-10 [FREE Full text] [doi: 10.3109/03009734.2015.1006347] [Medline: 25622619]
20. Dimitrov DV. Medical Internet of Things and Big Data in Healthcare. Healthc Inform Res 2016 Jul;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
21. Çolak A, Yılmaz C, Toprak B, Aktoğu S. Procalcitonin and CRP as Biomarkers in Discrimination of Community-acquired Pneumonia and Exacerbation of COPD. J Med Biochem 2017 Apr;36(2):122-126 [FREE Full text] [doi: 10.1515/jomb-2017-0011] [Medline: 28680355]
22. Park J, Park J. Pressed region integrated 3D paper-based microfluidic device that enables vertical flow multistep assays for the detection of C-reactive protein based on programmed reagent loading. Sensors and Actuators B: Chemical 2017 Jul;246:1049-1055. [doi: 10.1016/j.snb.2017.02.150]

Abbreviations

CRP: C-reactive protein
DPV: differential pulse voltammetry
ELISA: enzyme-linked immunosorbent assay
GLEIA: gold-linked electrochemical immunoassay
IoT: Internet of Things

Edited by G Eysenbach; submitted 22.03.20; peer-reviewed by B Shenker, I Gabashvili, J Carter; comments to author 10.04.20; revised version received 12.05.20; accepted 14.06.20; published 07.09.20.

Please cite as:
Gondoh-Noda Y, Kometani M, Nomura A, Aono D, Karashima S, Ushijima H, Tamiya E, Murayama T, Yoneda T
Feasibility of a Novel Mobile C-Reactive Protein–Testing Device Using Gold-Linked Electrochemical Immunoassay: Clinical Performance Study
JMIR Mhealth Uhealth 2020;8(9):e18782
URL: http://mhealth.jmir.org/2020/9/e18782/
doi:10.2196/18782
PMID:32894233

Original Paper

Identification of Type 2 Diabetes Management Mobile App Features and Engagement Strategies: Modified Delphi Approach

Hanan A Alenazi[1*], BSc, MPH; Amr Jamal[2,3*], MD, SBFM, ABFM, MRCGP, MBI, CPHIMS; Mohammed A Batais[2], MBBS, SBFM, ABFM

[1]National Health Information Center, Riyadh, Saudi Arabia

[2]Family and Community Medicine Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[3]Research Chair of Evidence-Based Healthcare and Knowledge Translation, King Saud University, Riyadh, Saudi Arabia

[*]these authors contributed equally

Corresponding Author:
Hanan A Alenazi, BSc, MPH
National Health Information Center
Saudi Health Council
Riyadh, 13315
Saudi Arabia
Phone: 966 502025959
Fax: 966 114579360
Email: H.Alenazi@shc.gov.sa

Abstract

Background: Diabetes is a significant public health issue. Saudi Arabia has the highest prevalence of type 2 diabetes mellitus (T2DM) in the Arab world. Currently, it affects 31.6% of the general population, and the prevalence of T2DM is predicted to rise to 45.36% by 2030. Mobile health (mHealth) offers improved and cost-effective care to people with T2DM. However, the efficiency of engagement strategies and features of this technology need to be reviewed and standardized according to stakeholder and expert perspectives.

Objective: The main objective of this study was to identify the most agreed-upon features for T2DM self-management mobile apps; the secondary objective was to identify the most agreed-upon strategies that prompt users to use these apps.

Methods: In this study, a 4-round modified Delphi method was applied by experts in the domain of diabetes care.

Results: In total, 11 experts with a mean age of 47.09 years (SD 11.70) consented to participate in the study. Overall, 36 app features were generated. The group of experts displayed weak agreement in their ranking of intervention components (Kendall W=0.275; *P*<.001). The top 5 features included insulin dose adjustment according to carbohydrate counting and blood glucose readings (5.36), alerting a caregiver of abnormal or critical readings (6.09), nutrition education (12.45), contacts for guidance if required (12.64), and offering patient-specific education tailored to the user's goals, needs, and blood glucose readings (12.90). In total, 21 engagement strategies were generated. Overall, the experts showed a moderate degree of consensus in their strategy rankings (Kendall W=0.454; *P*<.001). The top 5 engagement strategies included a user-friendly design (educational and age-appropriate design; 2.82), a free app (3.73), allowing the user to communicate or send information/data to a health care provider (HCP; 5.36), HCPs prescribing the mobile app in the clinic and asking about patients' app use compliance during clinical visits (6.91), and flexibility and customization (7.91).

Conclusions: This is the first study in the region consisting of a local panel of experts from the diabetes field gathering together. We used an iterative process to combine the experts' opinions into a group consensus. The results of this study could thus be useful for health app developers and HCPs and inform future decision making on the topic.

(JMIR Mhealth Uhealth 2020;8(9):e17083) doi:10.2196/17083

KEYWORDS

diabetes; mobile features; engagement strategies; mobile app; Delphi consensus

Introduction

Type 2 diabetes mellitus (T2DM) is one of the leading causes of disability and mortality worldwide, creating a substantial economic burden on health systems and social well-being [1,2]. In Saudi Arabia, studies evaluating T2DM prevalence showed a continuously increasing trend, reaching an age-adjusted prevalence of 31.6%, which is considered among the highest worldwide [3,4].

Diabetes is a chronic, complex syndrome requiring condition-related knowledge and self-management skills to optimize glycemic control and improve health outcomes. Mobile apps could be helpful tools for supporting chronic disease screening, enhancing the ability of people with diabetes to manage the disease, and ensuring their easy accessibility to health care services. These apps boost the opportunity to increase health care access for vulnerable populations. The apposite use of mobile health (mHealth) offers improved and cost-effective care to people with T2DM through improved diabetes self-management [5-7].

Globally, the annual growth in the penetration rate of internet and mobile phone use is estimated to be 13%. In Saudi Arabia, which has a total population of 33.25 million, 30.25 million (91%) are internet users, 25 million (75%) are active social media users, and there are 56.80 million (171%) mobile subscriptions [8]. However, the utility features of this technology need to be standardized and reviewed according to stakeholder and expert perspectives [9].

Several systematic reviews, meta-analyses, and meta-regression analyses found that digital or telemedicine interventions could be more effective in enhancing the outcomes of managing T2DM compared to standard care. The apps seem to increase the perception of self-care by contributing to better knowledge among people with T2DM. People with T2DM also become more confident in dealing with their illness, primarily due to a decrease in fears resulting from a lack of information; for example, they develop more confidence in how they should deal with potential hypoglycemic episodes [10-14]. Another systematic review, on the other hand, claims that there is limited evidence supporting the effectiveness of diabetes apps [15].

A systematic review that evaluated free mobile apps for control and self-management of type 1 diabetes mellitus (T1DM) available in 2015 found that 56 of these apps did not even meet the minimum requirements or work efficiently. While there was a significant number of available apps, the study results indicated that only 9 of the 65 reviewed mobile apps are adaptable and useful for self-management [16]. Most insulin dose calculator apps provide incorrect or inappropriate dosage recommendations, which puts patients at risk. Therefore, proficient health care providers (HCP) should be involved in the app design stage to address safety during self-management and health education [17]. There is a need for comprehensive, efficient, and flexible mobile apps for self-management of diabetes with more features to increase the number of long-term users and influence better self-management [16,17] and patient empowerment.

A study of several focus groups concerning patient perspectives and expectations of diabetes self-management and eHealth found that there were significant variations in the health care services experienced by patients with diabetes, and none of the evaluated services met all users' needs [18]. Patient groups seem to differ in expectations and needs concerning self-management and eHealth for self-management purposes. In a Dutch study, several participants worried about the implementation of eHealth being a consequence of budget cuts in care [19]. Therefore, we chose to engage experts and to use formal consensus methods that have been used to guide action in areas of research in which there is inconsistent or contradictory scientific evidence. Consensus techniques are an efficient and valid method for identifying and collecting information on a subject on which there is little evidence or agreement. Consensus techniques allow one to obtain quantitative estimates from qualitative strategies and to determine the degree of agreement among participants [20].

Delphi techniques were developed in the 1950s. They are a structured and multistage process in which a panel of experts is invited to be a part of a series of rounds to identify and achieve consensus on a specific issue. The consensus is sought through information feedback and iteration in the form of rounds and phases. The process is terminated when a consensus is reached. The anonymity offered by the Delphi method can reduce the inhibition that usually occurs during decision-making; individuals are more open in their answers when their answers are deidentified. The Delphi technique is useful when working with highly subjective elements in which it is difficult to determine their intrinsic value [21,22]. The Delphi technique is a unique method used to gain consensus from experts who hold varied viewpoints or professional experiences. This technique can be used to build agreement and allow groups to judge frameworks [23,24].

The main objective of this study was to identify the most agreed-upon features for T2DM self-management mobile apps. The secondary objective is to identify the most agreed-upon strategies that encourage users to use these apps. Our study aimed to determine an expert panel’s consensus opinion on two questions:

1. What are the diabetes management components that diabetes experts agree are most likely to be effective when delivered via a mobile app?
2. What are the engagement strategies that diabetes experts believe are the most likely to be effective for diabetes management when using a mobile app?

Methods

Study Design

A 4-round modified Delphi study was conducted to identify mobile app features and engagement strategies for T2DM self-management. There were three reasons for choosing the Delphi technique: (1) the Delphi technique has been useful for measuring group consensus concerning medical information technology when used in health care; (2) there was no established evidence regarding self-management of T2DM or

engagement strategies in the context of developing mobile apps, and (3) the outcomes of Delphi decisions focus on decision making in fields that are strongly susceptible to change. Individual decision makers have more influence than those with underlying rules [20,25]. We used supplemental techniques to ensure the validity of study outcomes by creating a steering committee that includes subject experts to oversee the design, execution, and analysis of all study phases. An agreement between the steering committee members was reached by discussion to approve the participant selection, consensus threshold, survey format and questions, and analysis process; in addition, they wrote down comments and discussed any open points during all 4 rounds. Thus, a modified Delphi technique was used to identify mobile app features and engagement strategies for T2DM self-management.

Participants

In total, 11 multidisciplinary experts were purposively selected from a range of scientific networks. Their professions included family physician, clinical informatician, diabetologist, clinical pharmacist, endocrinologist, exercise physiologist, nurse diabetes educator, health educator, clinical dietitian, consultant in medical education, and psychologist. Previous literature reports have indicated that the Delphi technique does not require a particular sample size. The minimum number of participants should be at least 7, although 10 to 20 are advisable [26,27]. A purposive sample is necessary to ensure the variability of the invited experts' background and experience. The research project steering committee used their expertise to judge the suitability of the invited experts. After identifying the experts, formal contact was established through email, phone, or an in-person meeting. The experts were also asked whether they could further recommend experts who could add value to the project, as a snowball sampling technique. All of the experts who were approached agreed to participate in the project. The experts were all working in Saudi Arabia. Detailed profiles of the respondents are provided in Table 1. The selection of experts participating in the study was based on their diverse professional backgrounds and experience related to the care of people with T2DM. Apart from the clinical informatician, knowledge related to diabetes apps was not one of the selection criteria. Having a similar professional background and experience of a preexisting participant was the only exclusion criteria for the study.

Table 1. Profiles of the expert panel members (N=11).

Professional title	Latest academic degree	Specialized fields	Years of experience in health care	Years of experience in diabetes care	Number of patients per week	Years of experience in health informatics research	Years of experience in diabetes research	Gender
Professor, public health researcher, consultant in family and community medicine	FRCGP	Diabetes prevention and management, primary care, public health, evaluation research, health services research, health promotion, health education, patient-centered care design	33	33	18	0	33	Male
Professor, consultant in family and community medicine, researcher	Diploma in Medical Education	Primary care research, patient-physician decision making, interprofessional education, public health research, medical education, knowledge translation, communication skills	32	32	14	2	32	Male
Assistant professor, researcher, consultant diabetologist	Fellowship in diabetes	Diabetes prevention and management (tertiary care), clinical trials, population health research, patient education, community participation	10	7	70	0	6	Male
Endocrinologist consultant, president of Saudi Society of Endocrinology and Metabolism	SBIM	Health services (not for profit), diabetes prevention and management (tertiary care), metabolic diseases	30	30	45	0	30	Male
Assistant professor, consultant, Director of King Abdullah Arabic Health Encyclopedia	Diploma in research methodology	Health care information technology, informatics, health care management, medical education, patient safety, clinical research, public health, telehealth	17	17	0	10	1	Male
Nurse educator, head of the education department	BSc	Patient education, diabetes prevention and self-management, clinical research	25	25	100	0	15	Female
Health education specialist	BSc	Health education, diabetes prevention and self-management, primary and tertiary care	5	3	50	0	1	Female
Consultant psychologist	Psychology fellowship	Psychology care, diabetes psychology prevention management care, clinical research, communication skills	10	5	10	0	5	Male
Senior clinical dietitian	MSc	Clinical research, diabetes nutrition prevention management care, primary and tertiary care	6	3	30	0	3	Female
Assistant professor, clinical pharmacist	PhD	Clinical trials, clinical pharmacist, patient education, patient adherence	19	7	8	0	4	Female
Assistant professor, consultant physiology	PhD	Diabetes prevention and management, primary and tertiary care	21	3	20	0	3	Male

Selection of the Appropriate Expert Panel

Experts were selected from hospitals, academic institutes, and associations in Saudi Arabia. Moreover, we also invited those who had participated in recent regional diabetes conferences. The inclusion criteria were comprised of 3 parameters: (1) clinicians or researchers with expertise in diabetes care, (2) who possessed at least 3 years of experience in the diabetes care field, and (3) who were willing to participate.

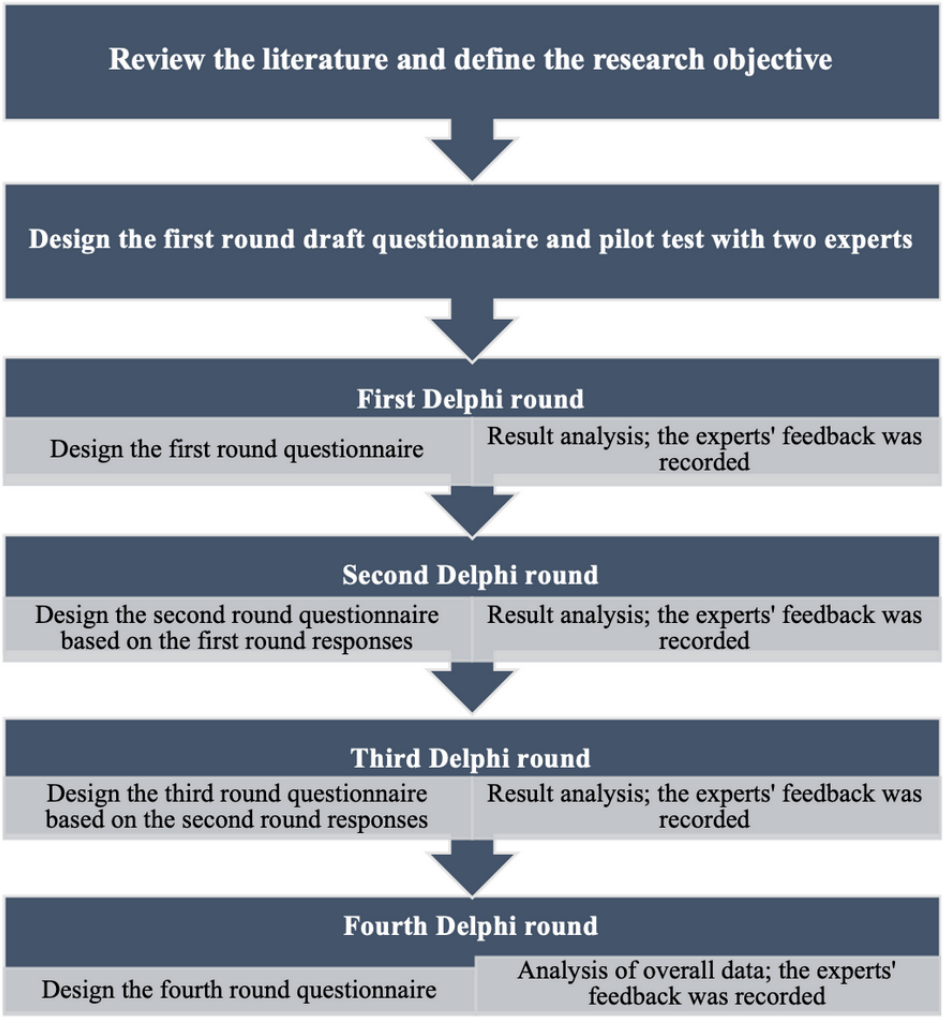
Steering Committee

This study included steering committee that was comprised of a diabetes educator; a health informatics specialist and consultant in family medicine and medical informatics; and a consultant in family medicine and diabetologist. The committee oversaw the design, execution, and analysis of all phases of this study. An agreement was reached by discussion to approve the participant selection, consensus threshold, survey format and questions, and analysis process.

Delphi Process

A minimum of 3 survey rounds were prospectively planned; rounds would continue until a consensus agreement was reached (Figure 1). The Delphi process started in November 2017 and ended in June 2018. Each round was conducted over 5 weeks, starting with the distribution of the materials. A reminder email was sent 1 week and 24 hours before the closure of the round, and repeated until the experts responded to the survey. The results generated from the previous round would be used in each subsequent round. For data collection in the first round, a questionnaire was sent via email. From the second round onward, an electronic survey instrument (SurveyMonkey) was used to enable efficient and timely data collection from the participants.

Figure 1. The modified Delphi process followed in this study.



Pilot Study

The primary draft survey was pilot tested by two experts with expertise in diabetes research who not included in the original study. Pilot testing determined the time it took to complete the survey and whether the survey wording was clear. After minor amendments arising from the pilot testing, the first round survey format was considered complete.

First Round

By using a survey with open-ended questions, the first round aimed to generate a list of items. Participants were asked to consider which items should be included when creating a mobile app designed to help people with diabetes self-manage their illness. The initial survey consisted of three sections: (1) a brief overview of the Delphi process; (2) demographic information; (3) two open-ended questions asking participants which items

should be included when designing the mobile app. The experts were invited to provide at least three responses to each question:

1. Which features or components of management care delivered by the mobile app would be most effective in helping patients with diabetes improve their health?
2. What are the best strategies or techniques that should be used to engage people with diabetes to use such a mobile app?

Each question was preceded with the following statement: "Please answer the questions below based on your clinical experience, knowledge of the guidelines, and relevant theories. Please provide at least 3 responses for each question."

Second Round

In the second round, a checklist was generated from the first round results submitted by the experts, and they were asked their opinion. Furthermore, they were also asked to accept, delete, or modify each item. Qualitative responses offered insight into the differences in experts' perspectives. The items had been modified via qualitative feedback.

Third Round

In the third round of the Delphi process, participants were invited to rate the importance of each item of the alphabetically ordered checklist generated from the first and second rounds. They were asked to rate their agreement with each of these techniques for the 3 different questions on a 5-point Likert-scale. The scale consisted of 5 responses: 1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, and 5=strongly agree, as well as the option, "I do not know."

Fourth Round

Experts were provided with the items listed alphabetically, alongside the mean of agreement ratings. Experts were asked to rank their responses from 1 (most likely to be the best) to n=the total number of items (least likely to be the best) for each question. At this stage, participants were only asked to rank responses for which there had been broad agreement in previous rounds. At this point, there was the option to make final comments, if any.

Data Analysis

Data Management

Quantitative data analyses were performed with SPSS (version 20.0; IBM Corp). The descriptive statistics for each item were calculated, including the mean agreement scores, standard deviation, and the Kendall W statistic which employs the chi-square test to test the independence of the component rankings.

Feedback

Based on Research and Development Corporation (RAND) recommendations regarding participant feedback, the findings from the previous round were compiled for Rounds 2, 3, and 4 [28]. The feedback report was provided to participants 2 weeks before the initiation of the next round. The feedback from the last round was offered to participants 1 month after the round ended.

First-Round Analysis

After the first round, the total number of completed questionnaires was 11. For each question, two individuals from the steering committee assessed, reviewed, and summarized the results of the first round independently. Similar ideas were clustered together into emerging components. Duplicate responses were merged. Responses that were not related to the study objectives, such as "consume high fiber diet through vegetable and fruit," were deemed not applicable and they were removed. The two reviewers met to discuss any differences arising from these independent analyses. Subsequently, the results were reviewed, and a discussion was held to validate the results and categorize them into domains.

Second-Round Analysis

The total number of completed questionnaires was 11. After all responses were received, the steering committee compared the responses and reviewed the proposals to generate the final list.

Third-Round Analysis

The total number of completed surveys was 11. Participants were asked to rate their agreement with the 36 mobile app features and 21 engagement strategies mentioned in the first and second rounds on a 5-point Likert scale. Descriptive statistics for each item were calculated, including the mean agreement scores and standard deviation.

Fourth-Round Analysis

In the final round, participants were asked to rank the mobile features from 1 (highest) to 36 (lowest). The value of the 36 features was determined in the first and second rounds. Moreover, participants were asked to rank each engagement strategy from 1 (highest) to 21 (lowest). The value of the 21 engagement strategies was determined in the first and second rounds. Descriptive statistics for each item were computed, including the mean ranking scores and standard deviation (SD). We used the Kendall W coefficient of concordance, which measures the extent to which judges agree on their rankings of items. The Kendall W statistic uses the chi-square test to investigate the independence of the components' rankings.

Ethical Considerations

Ethical approval to conduct this study was obtained from the King Saud University Institutional Review Board (number E-17-2608). All participants were asked to sign a consent form prior to starting the rounds. Experts were asked to create a list of the best features and engagement strategies, which were subsequently rated and ranked anonymously to avoid biases.

Results

Demographic Characteristics of Participants

A total of 11 experts consented to participate in the study. They had a mean age of 47.09 years (SD 11.70; range 28-62). In total, 4 rounds of Delphi exercises were completed over 6 months, and all of the expert participants (n=11) completed all 4 rounds.

Features

A total of 77 features were generated in response to the first question, “what are the diabetes management components that diabetes experts agree are most likely to be effective when delivered via an app,” with an average of 7 features mentioned per participant in the first and second rounds (Multimedia Appendix 1). Duplicate features were removed, and some were merged (n=36), and the resulting 36 features were categorized into 5 domains (Table 2). No further items were added after rounds 3 and 4 for the whole list of items. Of the 36 features, 11 had a high mean agreement rating (above 4.50). These features were “a reminder for the health care providers' appointments, screening, and routine lab tests” (4.73), “carbohydrate counter” (4.72), and “diet planning” (4.72). All of the next 6 features scored the same (4.64): “nutrition education,” “contacts for guidance if needed,” “offer patient-specific education tailored to the user’s goals, needs, and blood glucose readings,” “find the nearest urgent health care services/centers,” “database for local restaurants and stores providing diabetes-friendly foods,” and “providing Arabic and English versions”. Both “medication reminders by notifications” and “the synchronization with the system, syncing with electronic medical records/personal health records, syncing with glucometers, and continuous glucose monitors or insulin pumps” scored the same (4.54). Overall, the original group of experts displayed weak agreement (Kendall W=0.275) in their ranking of intervention components (χ^2_{35}=106.017, P<.001).

There were 5 top features: (1) adjusting insulin doses according to carbohydrate counting and blood glucose readings (5.36); (2) alert caregiver (such as by text message) for abnormal or critical readings (6.09); (3) nutrition education (12.45); (4) contacts for guidance if required (12.64); and (5) offering patient-specific education tailored to the user’s goals, needs, and blood glucose readings (12.90).

Table 2. Responses generated by the expert group concerning the features.

Rank	Responses generated[a]	Domain	Agreement rating[b]			Ranking score[c]	
			Mean (SD)	Mode	Agree: disagree[d]	Mean (SD)	Mode
1	Reminder for the health care providers' appointments, screening, routine lab tests, etc	Follow-up care	4.73 (0.47)	5	11:0	13.00 (7.58)	14
2	Carbohydrate counter and health diet planning	Nutrition and diet management	4.72 (0.47)	4	11:0	13.73 (7.72)	12
3	Nutrition education	Nutrition and diet management	4.64 (0.50)	5	11:0	12.45 (8.51)	3
4	Contacts for guidance if needed	Follow-up care	4.64 (0.50)	5	11:0	12.64 (12.22)	1
5	Offer patient-specific education tailored to the user's goals, needs, and blood glucose readings	Education	4.64 (0.50)	5	11:0	12.90 (11.54)	5
6	Find the nearest urgent health care services/centers	Follow-up care	4.64 (0.50)	5	11:0	15.18 (10.33)	12
7	Database for local restaurants and stores providing diabetes-friendly foods	Nutrition and diet management	4.64 (0.50)	5	11:0	19.45 (9.09)	25
8	Providing Arabic and English versions	Mobile design and features	4.64 (0.67)	5	10:1	21.36 (12.47)	29
9	Medication reminders by notifications	Medication	4.54 (0.69)	5	10:1	15.09 (7.28)	7
10	Syncing with electronic medical record/personal health record	Mobile design and features	4.54 (0.52)	5	11:0	24.82 (8.67)	32
11	Syncing with meters, continuous glucose monitors, or insulin pumps	Mobile design and features	4.54 (0.52)	5	11:0	27.18 (6.40)	26
12	Alert caregiver (eg, by SMS text message) for abnormal or critical readings	Home monitoring	4.45 (0.52)	4	11:0	6.09 (5.16)	2
13	Blood glucose monitoring diary; give results for averages in graphs	Home monitoring	4.45 (0.69)	5	10:1	14.64 (8.32)	9
14	Individualized exercise suggestions or prescriptions	Physical activity	4.45 (0.69)	5	10:1	14.82 (6.92)	13
15	Assessing medication adherence	Medication	4.45 (0.52)	4	11:0	15.73 (10.46)	6
16	Online consultation (communication and patient monitoring) by health care providers	Follow-up care	4.45 (0.69)	5	10:1	17.18 (8.54)	17
17	Prescription refill reminders	Follow-up care	4.45 (0.52)	4	11:0	17.73 (8.94)	16
18	Special occasion management during fasting, Hajj, sick days, and travel	Education	4.45 (0.69)	5	10:1	21.18 (9.63)	7
19	Medication reconciliation (a comprehensive list of medications should include all prescribed medications, herbals, vitamins, etc)	Medication	4.45 (0.69)	5	10:1	21.27 (8.95)	19
20	Create, view, and manage alerts for personalized target goals	Education	4.36 (0.50)	4	11:0	17.64 (9.32)	10
21	General education (such as foot care, wound care, dental care, psychiatric symptoms associated with diabetes, driving, how to use the glucometer)	Education	4.36 (0.81)	5	9:2	19.09 (10.23)	11
22	Database for traditional low-carbohydrate recipes and traditional food carbohydrates and calories	Nutrition and diet management	4.36 (0.92)	5	10:1	20.09 (10.14)	4
23	Sharing blood glucose readings with health care providers	Home monitoring	4.36 (0.50)	4	11:0	21.54 (8.64)	10
24	Activity sensor trackers (such as step counters) or integration with wearable trackers	Physical activity	4.27 (0.65)	4	10:1	17.18 (9.73)	6

Rank	Responses generated[a]	Domain	Agreement rating[b]			Ranking score[c]	
			Mean (SD)	Mode	Agree: disagree[d]	Mean (SD)	Mode
25	Reminders to check blood glucose levels and ketones if needed	Home monitoring	4.27 (0.47)	4	11:0	17.18 (9.58)	3
26	Food barcode scanner and meal picture detection to log food	Nutrition and diet management	4.27 (0.65)	4	10:1	19.00 (9.89)	6
27	Workouts and exercise demonstrations	Physical activity	4.27 (0.65)	4	10:1	24.54 (11.58)	33
28	Find the nearest health and fitness activities	Physical activity	4.18 (0.87)	4	10:1	24.36 (10.13)	34
29	Pill identifier, drug list name with detailed information	Medication	4.18 (1.33)	5	8:3	25.73 (11.05)	36
30	Integrating the app with common social media platforms	Social media and communication	4.18 (0.75)	4	9:2	27.45 (7.05)	26
31	A food diary to track meals	Nutrition and diet management	4.09 (0.70)	4	9:2	20.82 (8.33)	19
32	Using anxiety and depression scales and providing customized advice to see health care providers	Psychosocial care	4.09 (0.70)	4	9:2	30.00 (7.17)	29
33	Adjusting insulin doses according to carbohydrate counting and blood glucose readings	Medication	4.00 (0.77)	4	8:3	5.36 (6.47)	5
34	Allowing for chat services for communication between users	Social media and communication	4.00 (0.77)	4	10:1	21.54 (10.00)	8
35	Body measurement trackers	Nutrition and diet management	3.91 (0.94)	4	8:3	22.18 (7.92)	15
36	Alert caregiver (such as by SMS text message) for missed doses	Home monitoring	3.82 (0.98)	4	9:2	15.81 (11.98)	8

[a]The features are ordered in terms of the mean agreement score from round 3.

[b]Agreement rating, where 1=strongly disagree and 5=strongly agree.

[c]Ranking score, where 1=highest and 36=lowest.

[d]Agree: disagree is the ratio of “agree” and “strongly agree” to “neither,” “disagree,” and “strongly disagree,” which was used as an inclusion criterion for round 3.

Engagement Strategies

A total of 53 engagement strategies were generated in response to the second question, “what are the engagement strategies that diabetes experts believe are the most likely to be effective in diabetes management when using a mobile app.” There were an average of 3.5 strategies suggested per participant in the first and second rounds (Multimedia Appendix 2). After the merging of similar strategies, the number of engagement strategies was reduced to 21 (Table 3). The resultant 21 strategies were categorized into 5 domains. No further strategies were added after Rounds 3 and 4 for the whole list of strategies. Out of the 21 engagement strategies, 4 had a high mean of agreement rating (above 4.40): (1) the app should be a free app (4.64); (2) allow the user to communicate or send information/data to a health care provider (4.54); (3) a user-friendly design (such as educational and age-appropriate design; 4.45); and (4) flexibility and customization (4.45). Overall, the experts showed a moderate degree of consensus in their ranking of the strategies (Kendall W=0.454, χ^2_{20}=99.924, P<.001).

The top 5 engagement strategies included several parameters: (1) a user-friendly design (such as an educational and age-appropriate design; 2.82); (2) the app should be free (3.73): (3) allowing the user to communicate or send information/data to a health care provider (5.36); (4) health care providers prescribing the mobile app in the clinic and asking about the patients’ app use compliance during clinical visits (6.91), and (5) flexibility and customization (7.91).

Table 3. Responses generated by the expert group concerning engagement strategies.

Rank	Responses generated[a]	Domain	Agreement rating[b]			Ranking score[c]	
			Mean (SD)	Mode	Agree: disagree[d]	Mean (SD)	Mode
1	The app should be free	Cost	4.64 (0.67)	5	10:1	3.73 (3.35)	2
2	Allowing the user to communicate or send information/data to health care providers	Communication or support	4.54 (0.522)	5	11:0	5.36 (3.04)	8
3	A user-friendly design (such as educational and age-appropriate design)	Easy to use	4.45 (0.52)	4	11:0	2.82 (2.64)	1
4	Flexibility and customization	Easy to use	4.45 (0.52)	4	11:0	7.91 (6.92)	4
5	Cross-platform device syncing (sync between mobile and other devices)	Features	4.36 (0.67)	4	10:1	13.73 (5.60)	16
6	Allowing chat services so users can communicate and support each other	Communication or support	4.27 (0.47)	4	11:0	9.45 (5.28)	12
7	Taking feedback and adding new features	Communication or support	4.27 (0.47)	4	11:0	10.00 (4.86)	8
8	Having health coaches explain the usefulness of the app through health campaigns	Advertisement	4.27 (0.65)	4	10:1	10.36 (4.94)	10
9	Using accessibility features (such as text-to-speech for people with limited vision)	Features	4.27 (0.65)	4	10:1	11.00 (4.19)	12
10	Using varying teaching methods (such as audiovisual, illustration, alarm)	Communication or support	4.18 (0.75)	4	9:2	13.18 (5.88)	13
11	Allowing the users to share their progress with family, friends, others through integration with common social media platforms	Communication or support	4.18 (0.60)	4	10:1	13.45 (5.94)	17
12	Syncing with electronic medical records/personal health record	Features	4.18 (0.60)	4	10:1	13.45 (4.27)	20
13	Having the official websites and social media accounts of scientific associations and organizations recommend the use of the app	Advertisement	4.09 (0.54)	4	10:1	14.82 (4.12)	14
14	Health care providers prescribing the mobile app in the clinic and asking about the compliance of the patients to use the app during the clinical visit	Advertisement or follow-up	4.00 (1.26)	5	8:3	6.91 (3.64)	3
15	Providing educational posters in the patient waiting area with a QR code to download the app	Advertisement	4.00 (1.18)	4	9:2	8.91 (4.87)	6
16	Providing services from trustworthy or well-known health care providers	Advertisement	3.82 (1.40)	4	9:2	9.18 (5.15)	4
17	Connecting the app with popular activity tracking devices (such as smartwatches, bands)	Features	3.82 (1.54)	4	9:2	12.73 (5.31)	11
18	Using colloquial terms in the push notification and information	Features	3.82 (0.87)	4	8:3	17.64 (2.20)	18
19	Providing inspirational and motivational quotes	Communication or support	3.64 (1.36)	4	8:3	13.09 (4.70)	10
20	Rewarding the users by offering nonfinancial incentives (such as gamification)	Cost	3.27 (1.74)	4	7:4	14.82 (5.09)	11
21	Rewarding the users by offering financial incentives	Cost	3.18 (1.78)	3	5:6	18.45 (3.50)	21

[a]The strategies are ordered in terms of the mean agreement score from round 3.

[b]Agreement rating, where 1=strongly disagree and 5=strongly agree.

[c]Ranking score, where 1=highest and 21=lowest.

[d]Agree: disagree is the ratio of “agree” and “strongly agree” to “neither,” “disagree,” and “strongly disagree,” which was used as an inclusion criterion for round 3.

Discussion

Features

In this study, participating diabetes experts expected many features to be included in a health app for people with diabetes. However, their level of agreement was slightly low, likely due to the variety of the experts' backgrounds and the differing amounts of experience [29].

There is sufficient evidence from the literature supporting the effectiveness of the top features generated in this study, which are listed in descending order according to the level of rating agreement: (1) a reminder for the health care providers' appointments, screening, and routine lab tests, which has been mentioned in several studies as a common use of health technology [30] that significantly affects outcomes [31]; and (2) carbohydrate counter and health diet planning**,** similar to tracking diary, carbohydrate, and meal intake [32,33] were recommended in several studies by diabetes educators who expressed their enthusiasm for viewing detailed dietary macronutrients [29]. The policy statement by the International Diabetes Federation (IDF) Europe uses some categories to differentiate between mobile health apps directed at people with diabetes. The first differentiating category was tracking, logging, and making dietary recommendations [34]. A qualitative study for weight and health management design indicates that the core components of the app should include tailored meal recommendations and assistance with meal planning [35]. For the nutrition education feature, a review found that the most useful mobile health app would help the patient with lifestyle education, self-management, and designing a suitable diet [36]. The IDF Europe categories also mention that such apps should help people with diabetes make food choices, undertake carbohydrate and calorie counting, and calculate medication dosages (similar to an insulin bolus calculator) [34]. A nutrition education app effectively raises awareness in people with diabetes [37].

For the contacts for guidance feature, the current National Health Service (NHS) services offer telephone contact points for youth. In addition, services with transition nurses, coordinators, or support workers offer contact via SMS text messaging [38]. The Saudi Arabian Ministry of Health (MOH) has a Service Center (that can be reached by dialing 937), which offers 24/7 medical doctor consultations [39] and can be added to mobile apps in Saudi Arabia. Several health apps suggested contacting HCP without providing specific contact details, access to experts, or just-in-time resources that could provide this type of guidance [40]. Other apps provided contact details for the diabetes health care team [41,42] or emergency contact lists [38].

For the "offering patient-specific education tailored to the user's goals, their needs, and blood glucose readings" feature, end users appreciate that it saves time and provides instructions tailored to their specific condition [43]. Additionally, app designers should take into consideration context-sensitive details and condition information [33]. A study conducted to evaluate the effectiveness of diabetes apps discovered that 73% (11/15) of the apps on the market allowed the user to set goals and to visualize times they did not meet their goals, generally using a specific color to indicate hypoglycemia or hyperglycemia [32]. The app should facilitate goal setting (such as weight loss targets, fitness goals) [44] because diabetes is very individual; therefore, the app should be customizable [34]. Matching the participants' perspective, a study showed that participants wanted the medical app to provide information regarding health screening and functions that can assess their health; these must be personalized to them and trustworthy [45]. Targeted exploration of the literature found that action treatment plans and personalized health goals provided by HCP resulted in statistically significant outcomes in several studies [31].

Regarding the "finding the nearest urgent health care services/centers" feature, the app could detect the location of the user with the help of GPS and the global system for mobile communications (GSM) network, and thereby display information about the nearest medical centers. By clicking on a particular hospital, all information regarding that hospital could be provided; Furthermore, this is particularly advantageous for travelers as they could be connected with nearby health care centers if needed, which is a common concern for travelers with chronic diseases [46]. For the "database listing local restaurants and stores providing diabetes-friendly food" feature, a study modified a food database to provide symbolic food names and calorie information, including recipes, ingredients, and local food names [41], which can be added to the diabetes self-management health apps. Another study showed that 11 of 15 available diabetes management health apps on the market focused on tracking carbohydrate intake, while only 3 of 15 had a built-in food database [32]. One study showed that women with gestational diabetes liked that the illustrations of diet-related information could be customized to their culture [41].

For the "provision of Arabic and English versions" feature, a recent Chinese study showed that English- and Chinese-language diabetes self-management apps constitute more than 80% of the 2000 available diabetes apps. Furthermore, the Chinese and English apps have more downloads and are more comprehensive with regard to clinically relevant functions compared to diabetes apps in other languages [42]. A study showed that women with gestational diabetes had previously experienced challenges with care provision that involved the help of an interpreter; furthermore, translating medical terminology into other languages can be challenging [41]. To our knowledge, no study has evaluated the quantity and quality of Arabic diabetes self-management mobile apps.

For the "medication reminders by notifications" feature**,** this factor had a statistically significant relationship with the outcomes [31]. The IDF Europe has recommended that health apps should adopt SMS text messaging or push notifications for insulin injections. These reminders are essential as they could prevent hyperglycemia and long-term complications associated with prolonged uncontrolled diabetes [34].

For the "syncing with electronic medical records/personal health record" feature, a study showed that educators favored the integration of mobile phone–collected data into the health information system [29]. From the users' perspective, a study

showed that people with diabetes also wanted the app to be integrated into existing systems since it could help increase the dissemination of the app and improve app uptake [39]. For the "syncing with meters, continuous glucose monitors, or insulin pumps" feature, a study that surveyed youths with T1DM or their parents found that a glucometer-connected mobile app could increase an individual's engagement with other aspects of care (such as self-monitoring of blood glucose frequency) [47]. IDF Europe's position on mobile apps for diabetes mentioned that interoperability is an essential feature of such apps [34].

Engagement Strategies

Overall, there was a moderate degree of consensus among experts on their rankings of the engagement strategies that they believe are most likely to be effective in diabetes management when using a diabetic self-management health app. Strategies with the highest mean of agreement (ratings above 4.40) have already been addressed in the literature. Regarding the "app being free" feature, 36/48 (75%) of the included apps were free, and the average cost of the paid apps was Aus $4.37 (US $3.13) [48]. User ratings and prices are important factors determining app attractiveness, with variations across countries [49]. A study analyzing both free and paid apps found that more expensive and popular paid apps tended to have more drawbacks. This relationship between popularity and drawbacks emerges from the fact that more expensive and popular paid apps tend to have more functionality [50].

The importance of "allowing the user to communicate or send information/data to a health care provider" has been demonstrated at the patient and health care levels [34]. A thematic analysis

study focused on app-based interventions in managing chronic respiratory diseases, diabetes, and hypertension demonstrated the perceived ability of HCP-motivated patients and empowered them to properly self-manage their condition. The use of health technology in two-way communication between HCPs and patients proved to impact the patients' health outcomes [31].

Studies found that user-friendly design, simplicity, and intuition were vital aspects for engaging younger adults to use health apps [34,51-53]. IDF Europe's position on mobile apps for diabetes mentioned that determining the target audience is crucial for the uptake of an app [34].

Several studies have mentioned "flexibility and customization" as a requirement [29,54,55] and a key strategy to facilitate engagement with therapies. The ability to evaluate the app as a guest user, as well as the ability to modify the welcome message and color palate, are examples of customization and flexibility [56].

Strengths and Limitations

One of the strengths of our study was having two people with diabetes on the study panel of experts. Therefore, besides their professional experience, they provided input into how the app could contribute to their life as patients with diabetes. The number of experts included was within the recommended range. Moreover, all the rounds were summarized and discussed by the steering committee, who are experts in the field.

There were several potential limitations identified in this study. First, we recruited experts with various backgrounds, which could have caused the significant levels of disagreement in the ranking. At the same time, if we had selected other experts, that could have led to a different result. Furthermore, there was considerable overlap among many of the information items volunteered by the Delphi participants. The use of a Delphi approach for selecting intervention components is not guaranteed to result in the best choices, which might affect the results' reliability; this is a known limitation of the Delphi technique.

Another limitation is that our panel consisted only of experts. We did not include patients with diabetes, although two of the participants coincidentally had diabetes. To determine whether these same features and engagement strategies are desired by the end users, another validation study should be conducted among a population of people with diabetes to bridge the gap between the perspectives of experts and end users.

Conclusions

To our knowledge, this is the first study in the Middle East and North Africa that gathered a local panel of experts from the diabetes field and used an iterative process to combine the experts' opinions into a group consensus. Consensus agreement does not mean that all of the right answers have been found, but rather indicates that a level of participant agreement has been reached. The information items resulting from this modified Delphi survey represent the opinions of an expert panel.

This study allowed us to reach a consensus on several important questions related to diabetes management components that experts agreed were the most likely to be effective when delivered via a mobile app. Furthermore, it shed light on the engagement strategies that diabetes experts believed would be the most likely to be effective in diabetes management when using a mobile app. The results of this study could thus be useful for health authorities and HCPs for future decision making on this topic.

Recommendations

The 36 features and 21 engagement strategies identified in this study should guide developers considering mobile app development targeting people with diabetes and other similar chronic diseases. For further research, we recommend that the results of this study should be verified and validated by building a prototype app that includes the features and strategies described in this study and recruiting people with diabetes to test it. Moreover, it would be more interesting to know if any of the top 10 app downloads globally or regionally contain these features and which combinations are most commonly found. How easily are these features implemented in an app? Are there cultural contexts that make certain feature more appropriate? Are any of the features not currently found in any app? These and other questions remain to be answered.

Acknowledgments

We would like to thank all the experts who participated in this study. The project was supported by King Saud University, Deanship of Scientific Research, Research Chair of Evidence-Based Healthcare and Knowledge Translation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention components generated by the experts in the first round.
[DOCX File , 107 KB - mhealth_v8i9e17083_app1.docx]

Multimedia Appendix 2
Engagement strategies generated by the experts in the first round.
[DOCX File , 121 KB - mhealth_v8i9e17083_app2.docx]

References

1. Young-Hyman D, de Groot M, Hill-Briggs F, Gonzalez J, Hood K, Peyrot M. Response to Comments on Young-Hyman et al. Psychosocial Care for People With Diabetes: A Position Statement of the American Diabetes Association. Diabetes Care 2016;39:2126-2140. Diabetes Care 2017 Sep 22;40(9):e131-e132 [FREE Full text] [doi: 10.2337/dci16-0051] [Medline: 28830965]
2. Eliadarous H. Exploring the impact of diabetes in Sudan out-of-pocket expenditure and social consequences of diabetes on patients and their families. 2017. URL: https://openarchive.ki.se/xmlui/handle/10616/45956 [accessed 2020-08-30]
3. Al-Daghri NM, Al-Attas OS, Alokail MS, Alkharfy KM, Yousef M, Sabico SL, et al. Diabetes mellitus type 2 and other chronic non-communicable diseases in the central region, Saudi Arabia (Riyadh cohort 2): a decade of an epidemic. BMC Med 2011 Jun 20;9(1):76 [FREE Full text] [doi: 10.1186/1741-7015-9-76] [Medline: 21689399]
4. Alotaibi A, Perry L, Gholizadeh L, Al-Ganmi A. Incidence and prevalence rates of diabetes mellitus in Saudi Arabia: An overview. J Epidemiol Glob Health 2017 Dec;7(4):211-218 [FREE Full text] [doi: 10.1016/j.jegh.2017.10.001] [Medline: 29110860]
5. Heitkemper E, Mamykina L, Travers J, Smaldone A. Do health information technology self-management interventions improve glycemic control in medically underserved adults with diabetes? A systematic review and meta-analysis. J Am Med Inform Assoc 2017 Sep 01;24(5):1024-1035 [FREE Full text] [doi: 10.1093/jamia/ocx025] [Medline: 28379397]
6. Lian J, McGhee S, Chau J, Wong CK, Lam CL, Wong WC. Systematic review on the cost-effectiveness of self-management education programme for type 2 diabetes mellitus. Diabetes Res Clin Pract 2017 May;127:21-34. [doi: 10.1016/j.diabres.2017.02.021] [Medline: 28315575]
7. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews. PLoS One 2017 Mar 1;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
8. Global Digital Report 2018. We Are Social. 2018. URL: https://wearesocial.com/blog/2018/01/global-digital-report-2018 [accessed 2018-11-09]
9. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. Diabetes Care 2016 Nov 25;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
10. Tchero H, Kangambega P, Briatte C, Brunet-Houdard S, Retali G, Rusch E. Clinical Effectiveness of Telemedicine in Diabetes Mellitus: A Meta-Analysis of 42 Randomized Controlled Trials. Telemed J E Health 2019 Jul;25(7):569-583. [doi: 10.1089/tmj.2018.0128] [Medline: 30124394]
11. Su D, Zhou J, Kelley MS, Michaud TL, Siahpush M, Kim J, et al. Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials. Diabetes Res Clin Pract 2016 Jun;116:136-148. [doi: 10.1016/j.diabres.2016.04.019] [Medline: 27321329]
12. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of Mobile Apps to Support the Care of Patients With Diabetes Mellitus: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. JMIR mHealth uHealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
13. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of Digital Interventions for Improving Glycemic Control in Persons with Poorly Controlled Type 2 Diabetes: A Systematic Review, Meta-analysis, and Meta-regression Analysis. Diabetes Technol Ther 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
14. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018 Jun 26;18(1):495 [FREE Full text] [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]

15. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: A systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
16. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile Applications for Control and Self Management of Diabetes: A Systematic Review. J Med Syst 2016 Sep 13;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]
17. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015 May 06;13(1):106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
18. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital Health Interventions for Adults With Type 2 Diabetes: Qualitative Study of Patient Perspectives on Diabetes Self-Management Education and Support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
19. Huygens MWJ, Vermeulen J, Swinkels ICS, Friele RD, van Schayck OCP, de Witte LP. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Jul 08;16(1):232 [FREE Full text] [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]
20. Garnett C, Crane D, West R, Brown J, Michie S. Identification of Behavior Change Techniques and Engagement Strategies to Design a Smartphone App to Reduce Alcohol Consumption Using a Formal Consensus Method. JMIR mHealth uHealth 2015 Jun 29;3(2):e73 [FREE Full text] [doi: 10.2196/mhealth.3895] [Medline: 26123578]
21. Carratalá-Munuera MC, Gil-Guillen VF, Orozco-Beltran D, Navarro-Pérez J, Caballero-Martínez F, Alvarez-Guisasola F, "Integrated Management of Type 2 Diabetes Mellitus" Research Group. Barriers associated with poor control in Spanish diabetic patients. A consensus study. Int J Clin Pract 2013 Sep 12;67(9):888-894. [doi: 10.1111/ijcp.12160] [Medline: 23758484]
22. Durham J, Al-Baghdadi M, Baad-Hansen L, Breckons M, Goulet JP, Lobbezoo F, et al. Self-management programmes in temporomandibular disorders: results from an international Delphi process. J Oral Rehabil 2016 Dec 01;43(12):929-936. [doi: 10.1111/joor.12448] [Medline: 27727477]
23. Ogden SR, Culp WC, Villamaria FJ, Ball TR. Developing a Checklist: Consensus Via a Modified Delphi Technique. J Cardiothorac Vasc Anesth 2016 Aug;30(4):855-858. [doi: 10.1053/j.jvca.2016.02.022] [Medline: 27256448]
24. Paquette-Warren J, Tyler M, Fournie M, Harris SB. The Diabetes Evaluation Framework for Innovative National Evaluations (DEFINE): Construct and Content Validation Using a Modified Delphi Method. Can J Diabetes 2017 Jun;41(3):281-296 [FREE Full text] [doi: 10.1016/j.jcjd.2016.10.011] [Medline: 28279617]
25. Iljaž RJ, Meglič M, Svab I. Building consensus about eHealth in Slovene primary health care: Delphi study. BMC Med Inform Decis Mak 2011 Apr 18;11(1):25 [FREE Full text] [doi: 10.1186/1472-6947-11-25] [Medline: 21501514]
26. Magnan EM, Gittelson R, Bartels CM, Johnson HM, Pandhi N, Jacobs EA, et al. Establishing chronic condition concordance and discordance with diabetes: a Delphi study. BMC Fam Pract 2015 Mar 28;16(1):42 [FREE Full text] [doi: 10.1186/s12875-015-0253-6] [Medline: 25887080]
27. Powell C. The Delphi technique: myths and realities. J Adv Nurs 2003 Mar;41(4):376-382. [doi: 10.1046/j.1365-2648.2003.02537.x] [Medline: 12581103]
28. Fitch K, Bernstein S, Aguilar M, Burnand B, LaCalle J. The RAND/UCLA appropriateness method user's manual. RAND Corporation. 2001. URL: https://www.rand.org/pubs/monograph_reports/MR1269.html [accessed 2020-08-30]
29. Wang J, Chu C, Li C, Hayes L, Siminerio L. Diabetes Educators' Insights Regarding Connecting Mobile Phone- and Wearable Tracker-Collected Self-Monitoring Information to a Nationally-Used Electronic Health Record System for Diabetes Education: Descriptive Qualitative Study. JMIR mHealth uHealth 2018 Jul 26;6(7):e10206 [FREE Full text] [doi: 10.2196/10206] [Medline: 30049667]
30. Ramsey RR, Carmody JK, Holbein CE, Guilbert TW, Hommel KA. Examination of the uses, needs, and preferences for health technology use in adolescents with asthma. J Asthma 2019 Sep 12;56(9):964-972. [doi: 10.1080/02770903.2018.1514048] [Medline: 30207802]
31. Donevant S, Estrada R, Culley J, Habing B, Adams S. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc 2018 Oct 01;25(10):1407-1418 [FREE Full text] [doi: 10.1093/jamia/ocy104] [Medline: 30137383]
32. Martinez M, Park SB, Maison I, Mody V, Soh LS, Parihar HS. iOS Appstore-Based Phone Apps for Diabetes Management: Potential for Use in Medication Adherence. JMIR Diabetes 2017 Jul 11;2(2):e12 [FREE Full text] [doi: 10.2196/diabetes.6468] [Medline: 30291096]
33. Boyle L, Grainger R, Hall RM, Krebs JD. Use of and Beliefs About Mobile Phone Apps for Diabetes Self-Management: Surveys of People in a Hospital Diabetes Clinic and Diabetes Health Professionals in New Zealand. JMIR mHealth uHealth 2017 Jun 30;5(6):e85 [FREE Full text] [doi: 10.2196/mhealth.7263] [Medline: 28666975]
34. Rose KJ, Petrut C, L'Heveder R, de Sabata S. IDF Europe's position on mobile applications in diabetes. Diabetes Res Clin Pract 2019 Mar;149:39-46. [doi: 10.1016/j.diabres.2017.08.020] [Medline: 28951337]
35. Rivera J, McPherson AC, Hamilton J, Birken C, Coons M, Peters M, et al. User-Centered Design of a Mobile App for Weight and Health Management in Adolescents With Complex Health Needs: Qualitative Study. JMIR Form Res 2018 Apr 04;2(1):e7 [FREE Full text] [doi: 10.2196/formative.8248] [Medline: 30684409]

36. Szálka B, Kósa I, Vassányi I, Mák E. [Support of diabetes dietary management and self-management using mobile applications]. Orv Hetil 2016 Jul;157(29):1147-1153. [doi: 10.1556/650.2016.30458] [Medline: 27426463]
37. Gabrielli S, Dianti M, Maimone R, Betta M, Filippi L, Ghezzi M, et al. Design of a Mobile App for Nutrition Education (TreC-LifeStyle) and Formative Evaluation With Families of Overweight Children. JMIR mHealth uHealth 2017 Apr 13;5(4):e48 [FREE Full text] [doi: 10.2196/mhealth.7080] [Medline: 28408361]
38. Toft A, Taylor R, Claridge L, Clowes C, Ferguson J, Hind J, et al. The Experiences of Young Liver Patients Transferring From Children's to Adult Services and Their Support Needs for a Successful Transition. Prog Transplant 2018 Sep 13;28(3):244-249. [doi: 10.1177/1526924818781567] [Medline: 29898638]
39. 937 Services. Ministry of Health. URL: https://www.moh.gov.sa/en/937/Pages/default.aspx [accessed 2018-12-12] [WebCite Cache ID 74bxbccom]
40. 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]
41. Garnweidner-Holme L, Hoel Andersen T, Sando MW, Noll J, Lukasse M. Health Care Professionals' Attitudes Toward, and Experiences of Using, a Culture-Sensitive Smartphone App for Women with Gestational Diabetes Mellitus: Qualitative Study. JMIR mHealth uHealth 2018 May 14;6(5):e123 [FREE Full text] [doi: 10.2196/mhealth.9686] [Medline: 29759959]
42. Huang Z, Soljak M, Boehm BO, Car J. Clinical relevance of smartphone apps for diabetes management: A global overview. Diabetes Metab Res Rev 2018 May 15;34(4):e2990. [doi: 10.1002/dmrr.2990] [Medline: 29431916]
43. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR mHealth uHealth 2014 May 01;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
44. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR mHealth uHealth 2014 May 01;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
45. Zydney JM, Warner Z. Mobile apps for science learning: Review of research. Computers & Education 2016 Mar;94:1-17. [doi: 10.1016/j.compedu.2015.11.001]
46. Almobaideen W, Krayshan R, Allan M, Saadeh M. Internet of Things: Geographical Routing based on healthcare centers vicinity for mobile smart tourism destination. Technological Forecasting and Social Change 2017 Oct;123:342-350. [doi: 10.1016/j.techfore.2017.04.016]
47. Clements MA, Staggs VS. A Mobile App for Synchronizing Glucometer Data: Impact on Adherence and Glycemic Control Among Youths With Type 1 Diabetes in Routine Care. J Diabetes Sci Technol 2017 May;11(3):461-467 [FREE Full text] [doi: 10.1177/1932296817691302] [Medline: 28745097]
48. Nicholas J, Fogarty AS, Boydell K, Christensen H. The Reviews Are in: A Qualitative Content Analysis of Consumer Perspectives on Apps for Bipolar Disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
49. Kübler R, Pauwels K, Yildirim G, Fandrich T. App Popularity: Where in the World are Consumers Most Sensitive to Price and User Ratings? Journal of Marketing 2018 Sep;82(5):20-44. [doi: 10.1509/jm.16.0140]
50. Watanabe T, Akiyama M, Kanei F, Shioji E, Takata Y, Sun B, et al. Understanding the origins of mobile app vulnerabilities: A large-scale measurement study of free and paid apps. In: Proceedings of the 14th International Conference on Mining Software Repositories. 2017 Presented at: 14th International Conference on Mining Software Repositories; May 20-28 2017; Buenos Aires, Argentina p. 14-24. [doi: 10.1109/msr.2017.23]
51. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR mHealth uHealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
52. Teo CH, Ng CJ, White A. What Do Men Want from a Health Screening Mobile App? A Qualitative Study. PLoS One 2017 Jan 6;12(1):e0169435 [FREE Full text] [doi: 10.1371/journal.pone.0169435] [Medline: 28060953]
53. Salgado TM, Fedrigon A, Riccio Omichinski D, Meade MA, Farris KB. Identifying Medication Management Smartphone App Features Suitable for Young Adults With Developmental Disabilities: Delphi Consensus Study. JMIR mHealth uHealth 2018 May 23;6(5):e129 [FREE Full text] [doi: 10.2196/mhealth.9527] [Medline: 29792292]
54. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR mHealth uHealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
55. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
56. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth Codesign of a Mobile Phone App to Facilitate Self-Monitoring and Management of Mood Symptoms in Young People With Major Depression, Suicidal Ideation, and Self-Harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]

Abbreviations

IDF: International Diabetes Federation

mHealth: mobile health
T2DM: type 2 diabetes mellitus
T1DM: type 1 diabetes mellitus
HCP: health care provider

Edited by G Eysenbach; submitted 16.11.19; peer-reviewed by S Alomran, L Quinlan; comments to author 13.01.20; revised version received 25.04.20; accepted 03.06.20; published 11.09.20.

Please cite as:
Alenazi HA, Jamal A, Batais MA
Identification of Type 2 Diabetes Management Mobile App Features and Engagement Strategies: Modified Delphi Approach
JMIR Mhealth Uhealth 2020;8(9):e17083
URL: http://mhealth.jmir.org/2020/9/e17083/
doi:10.2196/17083
PMID:32678798

Review

Virtual Trauma-Focused Therapy for Military Members, Veterans, and Public Safety Personnel With Posttraumatic Stress Injury: Systematic Scoping Review

Chelsea Jones[1,2], MSc; Antonio Miguel-Cruz[1,3,4], PhD; Lorraine Smith-MacDonald[1], PhD; Emily Cruikshank[1,5], MEd; Delaram Baghoori[1,6], BSc; Avneet Kaur Chohan[1,3], BSc; Alexa Laidlaw[1,3], BSc; Allison White[1,3], BSW; Bo Cao[1,7], PhD; Vincent Agyapong[1,7], PhD; Lisa Burback[1,7], MD; Olga Winkler[1,7], MD; Phillip R Sevigny[1,5], PhD; Liz Dennett[1,8], MLiS; Martin Ferguson-Pell[1,6], PhD; Andrew Greenshaw[1,7], PhD; Suzette Brémault-Phillips[1,3], PhD

[1]Heroes in Mind Advocacy and Research Consortium, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[2]1 Field Ambulance Physical Rehabilitation Department, Canadian Armed Forces Health Services, Department of National Defense, Edmonton, AB, Canada

[3]Department of Occupational Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[4]Glenrose Rehabilitation Hospital Research Innovation and Technology (GRRIT), Glenrose Rehabilitation Hospital, Edmonton, AB, Canada

[5]Department of Educational Psychology, Faculty of Education, University of Alberta, Edmonton, AB, Canada

[6]Rehabilitation Science, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[7]Department of Psychiatry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

[8]Scott Health Sciences Library, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Chelsea Jones, MSc
Heroes in Mind Advocacy and Research Consortium
Faculty of Rehabilitation Medicine
University of Alberta
1-94 Corbett Hall
8205 - 114 Street
Edmonton, AB, T6G 2G4
Canada
Phone: 1 780 492 0404
Email: cweiman@ualberta.ca

Abstract

Background: A necessary shift from in-person to remote delivery of psychotherapy (eg, teletherapy, eHealth, videoconferencing) has occurred because of the COVID-19 pandemic. A corollary benefit is a potential fit in terms of the need for equitable and timely access to mental health services in remote and rural locations. Owing to COVID-19, there may be an increase in the demand for timely, virtual delivery of services among trauma-affected populations, including public safety personnel (PSP; eg, paramedics, police, fire, correctional officers), military members, and veterans. There is a lack of evidence on the question of whether digital delivery of trauma-therapies for military members, veterans, and PSP leads to similar outcomes to in-person delivery. Information on barriers and facilitators and recommendations regarding digital-delivery is also scarce.

Objective: This study aims to evaluate the scope and quality of peer-reviewed literature on psychotherapeutic digital health interventions delivered remotely to military members, veterans, and PSP and synthesize the knowledge of needs, gaps, barriers to, and facilitators for virtual assessment of and virtual interventions for posttraumatic stress injury.

Methods: Relevant studies were identified using MEDLINE (Medical Literature Analysis and Retrieval System Online), EMBASE (Excerpta Medica dataBASE), APA (American Psychological Association) PsycINFO, CINAHL (Cumulative Index of Nursing and Allied Health Literature) Plus with Full Text, and Military & Government Collection. For collation, analysis, summarizing, and reporting of results, we used the CASP (Critical Skills Appraisal Program) qualitative checklist, PEDro (Physiotherapy Evidence Database) scale, level of evidence hierarchy, PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews), and narrative synthesis.

Results: A total of 38 studies were included in this review. Evidence for the effectiveness of digital delivery of prolonged exposure therapy, cognitive processing therapy, behavioral activation treatment with therapeutic exposure to military members, veterans, and PSP was rated level 1a, whereas evidence for cognitive behavioral therapy was conflicting. The narrative synthesis indicated that virtual delivery of these therapies can be as effective as in-person delivery but may reduce stigma and cost while increasing access to therapy. Issues of risk, safety, potential harm (ie, suicidality, enabling avoidance), privacy, security, and the match among the therapist, modality, and patient warrant further consideration. There is a lack of studies on the influences of gender, racial, and cultural factors that may result in differential outcomes, preferences, and/or needs. An investigation into other therapies that may be suitable for digital delivery is needed.

Conclusions: Digital delivery of trauma therapies for military members, veterans, and PSP is a critical area for further research. Although promising evidence exists regarding the effectiveness of digital health within these populations, many questions remain, and a cautious approach to more widespread implementation is warranted.

(JMIR Mhealth Uhealth 2020;8(9):e22079) doi:10.2196/22079

KEYWORDS

trauma; mental health; telemedicine; therapy; rehabilitation; digital health; psychotherapy; military; veteran; first responder; public safety personnel; teletherapy; psychotherapy; telepsychiatry; mobile phone

Introduction

Background

Physical distancing during the COVID-19 pandemic has led to a rapid paradigm shift toward remote mental health service delivery and a surge in the use of digital health delivery (eg, teletherapy, telemedicine, eHealth, and mobile health) [1,2]. With the onset of COVID-19 and physical distancing rules for the containment of infection spread, supporting people at a physical distance with digital health delivery methods became necessary for accessing services, screening, assessment, and treatment [2,3]. Many aspects of legal, clinical, cultural, practical, and privacy and/or security issues remain to be addressed for delivering mental health services to trauma-affected populations, including public safety personnel (PSP; eg, border services, communications officials, correctional workers, firefighters, paramedics, police, etc), military members, and veterans [4]. mental health concerns and posttraumatic stress injuries (PTSIs) in these groups may be associated with professional service during, or exacerbated by, the COVID-19 pandemic. This review aims to systematically identify the scope of what is known in this context and to summarize current evidence supporting the use of digital health with military members, veterans, and PSP, together with a discussion of barriers and facilitators.

Military members and PSP are frequently exposed to potentially high stress and traumatic experiences in the course of service [5]. Such exposures can impact their mental and psychosocial health and result in PTSIs, a range of challenges from posttraumatic stress disorder (PTSD) to symptom clusters that may not meet diagnostic criteria but interfere with daily functioning in social, work, or family activities [4]. PTSD is the most common PTSI experienced by military members and veterans [6-8] and remains to be the predominant focus of most military and veteran health research and care [9-11]. Characterized by intrusion symptoms, avoidance, changes in cognition and mood, and changes in arousal and reactivity [12], PTSD has historically been difficult to treat because of the variety of associated symptoms. Isolated or cumulative traumatic experiences can also cause long-term psychological and spiritual struggles, including depression, anxiety, and moral injury [9,11,13,14]. Moral injury, a separate trauma syndrome that results from exposure to morally injurious experiences such as witnessing or participating in acts that transgress personal morals and values [14], is potentially a key PTSD comorbidity [15].

The incidence of PTSIs among military members, veterans, and PSP varies: within military and veteran populations globally, PTSD prevalence is persistent, complex, and may increase over time [16,17]. Among US military members deployed during the War on Terror, PTSD prevalence estimates reached 19% [17] compared with 5.3% for Canadians [18], 2.7% to 4% in UK military members [19], and 3% for military members from the Netherlands [20]. In 2010, the rate of PTSD among the Australian Defense Force was reported to be 8.3% [21]. A recent meta-analysis reported that overall rates remained high (approximately 23%) for post–9/11 US veterans [22], and PTSD increased to 16% for Canadian veterans [23]. Global PTSI studies among PSP also demonstrate elevated prevalence, severity, and complexity. Prevalence of PTSIs among PSP in a recent Canadian study was reported as follows: municipal police, 36.7%; firefighters, 34.1%; Royal Canadian Mounted Police, 50.2%; and paramedical staff, 49.1% [24].

Multiple gold-standard frontline psychotherapeutic interventions have been used to treat PTSIs in these populations, including prolonged exposure therapy (PE), cognitive processing therapy (CPT), cognitive behavioral therapy (CBT), and eye movement desensitization reprocessing (EMDR). PE and CPT have been the main research focus over the last 25 years and appear to have the greatest efficacy for PTSD [25-27]. Although mental health interventions have been predominantly delivered in-person, the pandemic and geographical barriers impede in-person access to assessment and interventions. On the basis of recent evidence-based publications, some of the main barriers to veterans and military members seeking treatment include concerns about the specific treatment itself (ie, CBT, PE, CPT), mental health stigma from self and others, and technological logistics, including internet quality, familiarity with software, and competence with technology [28]. Furthermore, reduced access to mental health clinicians who are adequately trained in the remote delivery of therapy and individual factors, such

as unfavorable attitudes toward interventions, may also be significant barriers to receiving and benefiting from digital health treatments [29]. There is very limited access to specialized mental health services in active theaters of military operations, in remote locations, and for those living in rural communities [30]. The new reality of COVID-19 physical distancing requirements adds to existing barriers for those who may require medical treatment and services for mental and physical health. Increasing access to and understanding the effectiveness and limitations of virtual care is now essential.

Digital health (eg, teletherapy, telemedicine, eHealth, and mobile health) may offer military members, veterans, and PSP alternative access to mental health services and therapies in a timely manner. Current literature identifies accepted conventional benefits associated with in-person delivery of gold-standard frontline psychotherapies, including flexible delivery times, physical privacy that enables avoidance of stigma, enhanced self-efficacy, and minimized negative attitudes toward mental health interventions [29,31]. In contrast, potential concerns about the provision of therapy via digital health include potential practical issues with technology interactions, client willingness to engage in telehealth, privacy, and safety concerns (eg, how adverse events will be handled remotely), and mental health clinician attitudes.

Although there is evidence regarding the general use of digital health, significant knowledge gaps exist regarding its use in this context, including an understanding of needs, barriers, and facilitators of the use, uptake, and sustainability of digital health; technological issues that may impede the use of digital health solutions (eg, feasibility, logistics, security, firewalls, compatibility, and policies in military and PSP organizations) and technology acceptance by PSP, military members, veterans, and mental health clinicians; and evidence of clinical effectiveness of remote digital assessment and treatment of PTSIs. There is also, as yet, a general lack of published work focusing on knowledge dissemination and implementation plans that will ensure that timely, accessible, and relevant evidence is available to decision makers who may consider the adoption of virtual delivery of assessment and interventions. This review explores the literature regarding digitally delivered psychotherapeutic interventions for military members, veterans, and PSP, along with barriers, facilitators, and recommendations for its use. This study contributes uniquely to mental health response to COVID-19 by summarizing the evidence for realistic digital health solutions for the delivery of mental health services for military members, veterans, and PSP with PTSIs for whom mental health challenges may be associated with or exacerbated by the COVID-19 pandemic.

The research questions guiding this systematic scoping review are as follows:

1. What is the quality of the existing literature addressing digital health delivery of gold-standard psychotherapeutic trauma interventions for military members, veterans, and PSP with trauma-related mental illness?
2. What evidence exists on the efficacy of digital health delivery of gold-standard psychotherapeutic trauma interventions for military members, veterans, and PSP with trauma-related mental illness compared with regular in-person intervention delivery?
3. What are the facilitators, barriers, themes, clinical recommendations, considerations, and knowledge gaps in the current peer-reviewed, evidence-based literature regarding digital health delivery of PTSI-trauma interventions for military members, veterans, and PSP with trauma-related mental illness?

Objectives

This scoping review aims to (1) systematically evaluate the quality of the existing quantitative, qualitative, and mixed methods peer-reviewed literature on digital health interventions for mental health with military member, veteran, and PSP populations and (2) synthesize the knowledge of needs, gaps, barriers, and facilitators for digital health delivery of PTSI assessment and interventions based on the existing literature.

Methods

Identification of Relevant Studies

This scoping review employed the following overarching steps: (1) formulation of the research questions based on PICO (Population, Intervention, Comparison, and Outcome) guidelines, (2) identification of relevant studies, (3) selection of studies, (4) charting of the data, and (5) collation, analysis, summarization, and reporting of results [32]. This scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting guidelines [33].

Information Sources and Search Strategy

The research team developed a search strategy based on specific inclusion and exclusion criteria. It included the following databases: MEDLINE (Medical Literature Analysis and Retrieval System Online; Ovid MEDLINE ALL), EMBASE (Excerpta Medica dataBASE; Ovid interface), APA (American Psychological Association) PsycINFO (Ovid interface), CINAHL (Cumulative Index of Nursing and Allied Health Literature) Plus with Full Text (EBSCOhost interface), and Military & Government Collection (EBSCOhost interface). The search consisted of an extensive list of keywords and subject headings covering 4 concepts: (1) traumatized individuals, (2) military or rescue personnel, (3) person therapy conducted remotely through technology (telephone, videoconferencing, or online), and (4) specific trauma-informed therapies (Multimedia Appendix 1). The 4 concepts were then combined with the Boolean AND. The search was limited to articles published from 2010 onward to include only current technology and therapeutic techniques. Studies were also limited to peer-reviewed articles in the English language. Editorials and other nonresearch articles were removed where possible. The full search strategy is available in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Articles selected for inclusion in this scoping review addressed military members, veterans, and/or PSP who had a primary diagnosis of PTSD and/or a trauma-related mental health disorder. The intervention modality in outcome studies, or being

targeted for validity or reliability studies or reviews, was limited to psychotherapeutic treatments administered via a remote platform. Psychotherapeutic interventions most strongly recommended for military members, veterans, and PSP include CPT, CBT, PE, and EMDR [34-38]. Other trauma-informed therapies were also included in the search, including motivational interviewing, adaptive disclosure therapy (ADT), and accelerated resolution therapy (ART), although the search terms would include results containing other trauma therapies [39,40]. Remote psychotherapeutic interventions were limited to those delivered via a remote person-to-person interaction (ie, the participant and clinician were in separate physical locations during intervention delivery). The delivery platforms included videoconferencing, text, app-based, virtual reality, or telephone communication, provided interaction between the participant and clinician was performed in real time and not prerecorded or automated. Both individual and group interventions were included.

If the published work included participants with comorbid conditions, such as other mental health disorders, disrupted sleep, chronic pain, substance use disorder, or traumatic brain injury, it was included if the additional conditions were secondary to the trauma-induced diagnosis and not the primary focus of the specific research study.

All articles included in the data set were peer reviewed and were quantitative, qualitative, mixed methods, and meta-analyses, regardless of positive, negative, or neutral findings. Articles were excluded from the review if they did not meet the inclusion criteria, were not peer reviewed, or were commentaries, editorials, or gray literature such as nonpublished graduate student theses. Studies that exclusively addressed civilians, such as those that focused on family members of military, veteran, and PSP populations only, were also excluded. Studies of interventions that were not specific to trauma, such as dialectical behavioral therapy, were excluded from the final selection of articles.

Selection of Studies

The study selection phase followed a variation of the procedures used by Neubauer et al [41] and Miguel Cruz et al [42]. First, a member of the research team exported all of the identified studies to the reference manager software EndNote X9.3.2 (Thomson Reuters) [43]. After deduplication, the references were imported to Covidence [44]. Second, before the title and abstract evaluation phase, members of the research team were trained in applying the inclusion and exclusion criteria (calibration phase). Then, 2 pairs of independent researchers evaluated the titles and abstracts of the remaining studies and compared them with the inclusion and exclusion criteria. Differences between the 2 pairs of independent researchers regarding the decision of whether or not to include a study in the next phase were addressed at subsequent meetings. During the full paper reading phase, 2 researchers reviewed the full texts of the selected studies. Each researcher independently assessed the studies to determine their suitability for inclusion in the data extraction phase. Inclusion or exclusion into the data set for analysis required consensus.

Charting of the Data: Data Extraction Process

During this phase, the researchers completed the data extraction of the final selected papers and met regularly to reconcile the differences that arose through discussion. In case of any disagreement, one of the researchers acted as a third rater. In addition, the co–principal investigators validated the data extracted from the studies. In each selected study, the research team extracted data according to the following domains: population (medical condition, age, branch of the military, race or ethnicity, sample size [N], and mean age [SD] in years), study features (design, outcome variables, and assessment tools), intervention (whether the intervention was provided via remote access, mode of delivery, therapy or care delivered, or in a group or 1:1 session), clinical assessment, clinical outcome measures, clinical outcomes, assessment of technology usability, technology outcome measures, technology, use outcomes, duration, and data analysis strategies.

Analysis, Summarization, and Reporting

Data Analysis

All data were analyzed and validated by at least two team members involved in the analysis. The research team met regularly to discuss data extraction, analysis, and synthesis, which were iterative and, in some cases, concurrent. Any discrepancies in the analysis of quantitative or qualitative data were resolved through discussion. This nonlinear process guaranteed rigor and internal validity.

A thematic analysis and narrative synthesis were used to qualitatively analyze the studies and compile the results. Data immersion occurred before the commencement of the analysis and coding process. The thematic analysis involved examining the text in detail to identify recurring patterns (*themes*) through both inductive and deductive reasoning [45]. The framework by Braun and Clarke [45] for qualitative thematic analysis guided the inductive analysis such that no pre-existing coding frame was imposed on the studies. The deductive analysis was guided by the research questions, particularly barriers, facilitators, and recommendations associated with the use of digital health [45]. Once the coding structure was developed, the first round of coding was completed by 2 team members. Open codes were later combined into preliminary patterns focusing on similarities and differences within and between studies. More abstract concepts were assigned to broader categories of themes and verified through key quotes. The data were then reviewed and recoded twice. Team members continually compared themes and resolved their differences through discussion. Following the thematic analysis, a narrative synthesis was conducted to organize, describe, explore, interpret, and fundamentally tell the story of the analysis [46,47].

Quality of the Evidence

The 38 selected studies were further analyzed to determine the quality of the evidence. The 3 tools utilized for this step were the PEDro (Physiotherapy Evidence Database) scale [48], CASP (Critical Appraisal Skills Program) qualitative checklist [49], and the levels of evidence hierarchy. The levels of evidence used to summarize the findings are based on the levels of evidence developed by Straus et al [50,51]. The CASP

qualitative checklist can be utilized as a tool to evaluate whether qualitative literature is valuable to the research topic of interest. The PEDro scale can assist researchers in rapidly identifying which randomized controlled trials (RCTs) are likely to be internally valid and could have sufficient statistical information to make their results interpretable [48]. For RCTs, studies scoring 9 to 10 on the PEDro scale are considered to be of *excellent* methodological quality. Studies with PEDro scores ranging from 6 to 8 are considered to be of *good* quality, whereas studies scoring 4 or 5 were of *fair* quality. Studies that scored below 4 were considered to be of *poor* quality [50].

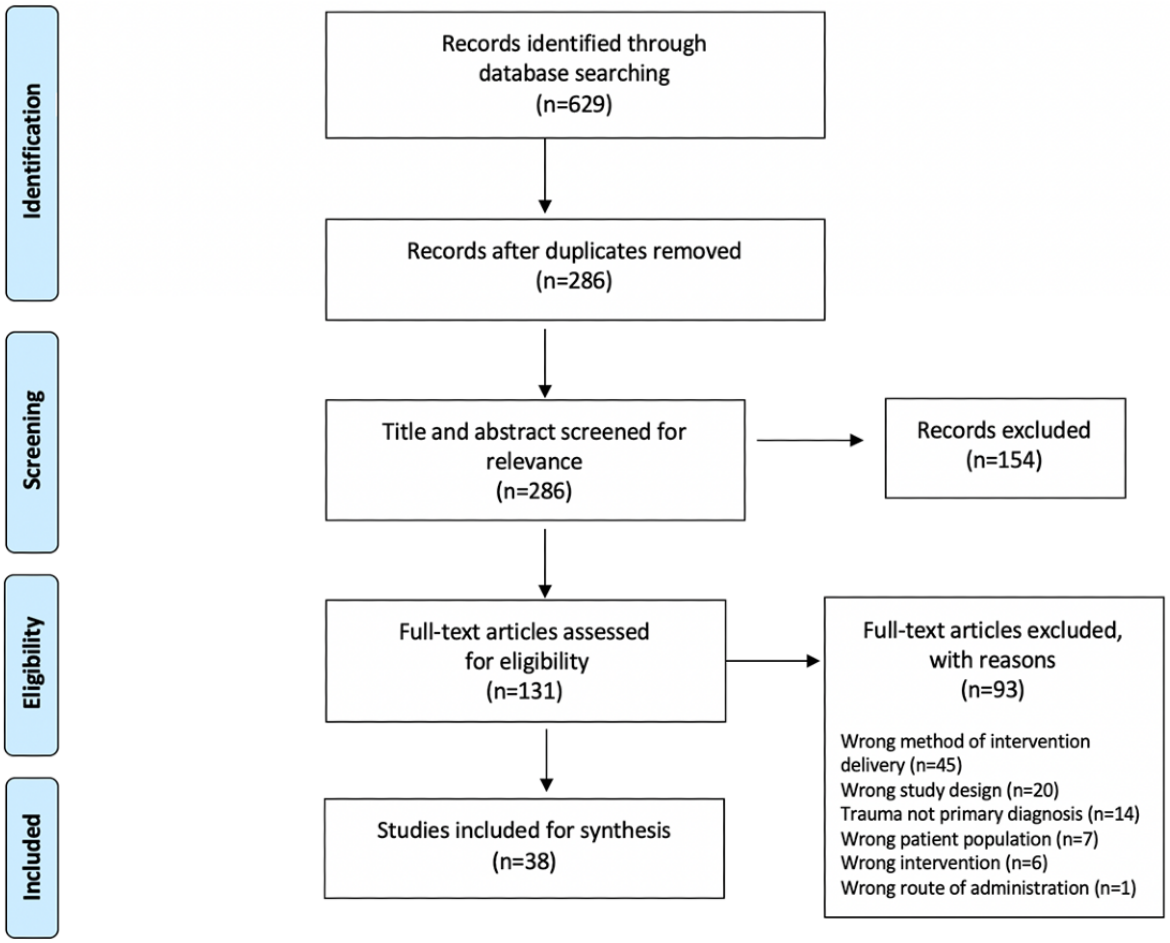
At least two researchers evaluated each of the included studies with the appropriate tool, with the PEDro scale being applied to RCTs (n=26) and the CASP qualitative checklist being applied to studies with a qualitative component (n=4). Each study with a quantitative component with outcome variables was then categorized based on the trauma therapy intervention used in the study and assigned a grade based on the levels of evidence [50,51].

Results

Search Results

The search strategy yielded 629 articles (see the PRISMA diagram in Figure 1). After deduplication, 286 titles and abstracts were screened. A total of 131 full-text documents were reviewed, with 93 being excluded (Figure 1) for reasons of including a therapeutic intervention not specific to traumatic stress-related disorders, not involving military members, veterans, or PSP, not being peer reviewed, the diagnosis not being related to trauma, or the administration of the therapeutic intervention not being via digital health. The remaining 38 studies were included in the review. Results of the descriptive analysis are displayed in Multimedia Appendix 2. In Multimedia Appendix 3 [25,28,52-87], a summary of the results of each included study is displayed.

Figure 1. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for Scoping Reviews chart of the scoping review study identification, selection, exclusion, and inclusion.



Evidence Appraisal Results

The RCTs (n=29) varied in quality, with PEDro scores ranging from 4 to 9. The CASP qualitative scale found all studies with a qualitative component (n=4) to be valuable and to have contributed to the overall evidence on the topic. Levels of evidence (Table 1) ranged from 1a (PE, CPT, and behavioral activation [BA] and therapeutic exposure [BA-TE]) to 4 BA. A level 1a rating indicates strong evidence and is achieved when 2 or more RCTs of high quality (PEDro ≥6) demonstrate similar findings, and a level 4 rating indicates that findings are supported by at least one study of pre- and posttest, posttest, case series, or single-subject intervention design [50]. The results for CBT were conflicting (Table 1).

Table 1. Level of evidence hierarchy for digital health delivery of trauma therapy interventions utilized with military members, veterans, and public safety personnel.

Treatment	Conclusion of level of evidence	Study
Prolonged exposure therapy	There is level 1a evidence that prolonged exposure therapy delivered via videoconferencing significantly reduces PTSD[a] symptoms in veterans and military members with PTSD	• Yuen et al, 2015 [25] • Wierwille et al, 2016 [53] • Tuerk et al, 2010 [58] • Morland et al, 2019 [83] • Jaconis et al, 2017 [70] • Hernandez-Tejada et al, 2017 [72] • Gros et al, 2011 [74] • Gros et al, 2018 [75] • Franklin et al, 2017 [77] • Pelton et al, 2015 [80] • Olden et al, 2017 [81] • Acierno et al, 2017 [86]
Cognitive processing therapy	There is level 1a evidence that cognitive processing therapy delivered via videoconferencing significantly reduces PTSD symptoms in veterans with PTSD	• Wierwille et al, 2016 [53] • Wells et al, 2019 [56] • Morland et al, 2015 [63] • Morland et al, 2014 [64] • Morland et al, 2011 [66] • Maieritsch et al, 2016 [67] • Fortney et al, 2015 [78] • Murphy and Turgoose, 2019 [82]
Cognitive behavioral therapy	There is conflicting evidence that cognitive behavioral therapy delivered via videoconferencing or telephone significantly reduces PTSD symptoms in veterans and military members with PTSD	• Ziemba et al, 2014 [52] • Stecker et al, 2014 [28] • Stecker et al, 2016 [62] • Gallegos et al, 2016 [76] • Trahan et al, 2016 [59]
Behavioral activation and therapeutic exposure	There is level 1a evidence that behavioral activation and therapeutic exposure delivered via home-based videoconferencing significantly reduces PTSD symptoms in veterans with PTSD	• Acierno et al, 2016 [87] • Strachan et al, 2012 [60]
Behavioral activation treatment for PTSD	There is level 4 evidence that behavioral activation for PTSD delivered via clinic-based videoconferencing significantly reduces PTSD symptoms and depressive symptoms in military members with PTSD	• Luxton et al, 2015 [68]

[a]PTSD: posttraumatic stress disorder.

Qualitative Analysis Results

The narrative synthesis and qualitative thematic analysis yielded a number of emerging themes related to the efficacy, clinical utility, ethics, accessibility, facilitators and barriers regarding the remote delivery of trauma-focused therapies. In addition, a number of clinical recommendations have emerged that complement the findings of the quantitative analysis and rating of evidence.

Facilitators

A number of facilitators were identified regarding the virtual delivery of trauma therapies to military members, veterans, and PSP (Table 2). These included (1) the convenience of accessing teletherapy, particularly for clients in rural and remote areas [79,85,87], (2) the comfort of participating in therapy from the client's home [57,59,85,87] resulting in less stress [55,58] and stigma, [86,87], (3) the efficaciousness of several different evidence-based PTSD treatment modalities delivered using digital health, including PE, CPT, CBT, and BA-TE [25,53,56,77], and (4) the ability to see a therapist from a central health clinic or the location of their choosing reduced travel time and transportation and missed work costs [25,55,82]. Many clients found that participating in teletherapy provided the same opportunities for relationship building with the therapist as in-person treatment [52,81]. From the mental health clinician perspective, the same clinical skills and safety protocols required for handling increased emotional responses and symptom emergence in-person were noted to be used in a remote delivery context [58].

Table 2. Summary of facilitators from all studies included in the review.

Themes and subthemes	Findings	Illustrative quotes
1. Facilitators		
1.1 Participants		
1.1.1 Availability and accessibility of services	Access. Teletherapy may be beneficial for individuals who live in rural areas, as they are more accessible than in-person services. This benefit may be increased if the internet or electronic devices are provided to the clients	• "Service members who are living in geographically remote locations or in areas that have a shortage of specialty healthcare professionals may especially benefit from Home-Based Tele Mental Health options." [88]
1.1.2 Ethnic background and sex	Ethnicity and sex. Black veterans seem to be more likely to seek out services, whether through telehealth or in-person. Additionally, female military veterans may be more open to using teletherapy compared to in-person treatment	• "Blacks overall were found to be more than 2 times as likely to seek treatment as White participants." [62] • "...telehealth may help to overcome unique barriers experienced by female Veterans seeking care in a traditionally male-dominated health care system. Adoption of telehealth technologies may be particularly useful as the VAMC continues its efforts to expand services sensitive to the experiences of female Veterans including an expanded awareness and focus on providing MST related services." [70]
1.1.3 Rapport/trust building in therapy	Rapport. The ability to build rapport and develop a strong therapeutic alliance is possible with teletherapy and has been demonstrated in a number of studies	• "Participants reported high levels of therapeutic alliance with their therapist throughout the treatment." [81]
1.1.4 Participant environment	Home environment. Having the client participate in therapy from their own home allows them to feel comfortable and engage more easily in the therapeutic process. This may be especially helpful for clients who have experienced military sexual trauma	• "Participants mentioned that being able to do therapy in their own environment helped them to relax and engage better than if they had had to go somewhere unfamiliar." [85]
1.1.5 Use or uptake of the therapy	Uptake. Evidence seems to point to the comfort of clients with the use of teletherapy, belief in its effectiveness, and a willingness to use it again	• "...throughout the duration of treatment, the majority of participants reported that they would be willing to use telehealth-delivered treatment again." [81] • "Participants also endorsed high expectations that the intervention would be helpful throughout the course of treatment." [81]
1.2 Technology		
1.2.1 Stigma associated with therapy	Stigma. There seems to be a reduced amount of stigma surrounding teletherapy compared to in-person due to issues of privacy	• "The Advantage of Home-Based Teletherapy include reduced stigma (eg, patients do not need to visit a mental health care facility)..." [87]
1.2.2 Availability or accessibility or cost-effectiveness of services	Cost. Home-based teletherapy may be less expensive than classic in-person therapy, thereby making it more accessible to clients and decreasing the cost (transportation costs, travel time, and missed work). Clients also seem to appreciate the flexibility of teletherapy in terms of where and when they can access treatment	• "It is more convenient and it is not like waiting at the office knowing you just have 1 hour to talk." [59]
1.3 Ethical		
1.3.1 Privacy	Security. Some clients see teletherapy as more private and secure because they can access it in their homes. This is especially apparent in smaller, tight-knit communities	• "Moreover, some service members may be drawn to HBTmental health because of the privacy it offers to those who are concerned about stigma associated with seeking mental health treatment." [88]

Themes and subthemes	Findings	Illustrative quotes
1.3.2 Safety or risk	Stress. Teletherapy is perceived by some clients as less stressful than in-person therapy. The same clinical skills can be used during remote as in-person delivery for handling heightened emotional responses and symptoms	• "Titrating of emotional reactions and patient engagement in traumatic memories, normally including anxiety, increased psychomotor activity, crying, and reexperiencing symptoms, were all handled adequately with the same protocol and clinical skills employed for in-person PE" [58]
1.4 Clinical utility		
1.4.1 Effectiveness of different types of therapy when delivered via teletherapy	Modalities. Several different evidence-based therapies that have been shown to be effective for teletherapy including prolonged exposure therapy, cognitive processing therapy, and cognitive behavioral therapy	• "Use of clinical video teleconferencing services to provide evidence-based treatment to Veterans with posttraumatic stress disorder (PTSD) was found to be as effective as face-to-face treatment provision without negatively impacting therapeutic process measures." [64]
1.4.2 Therapy dropout rates in digital health	Dropout. Therapy dropout rates, reasons, and patterns are similar between therapy delivered in-person and remote delivery or therapy dropout rates are dependent more on comorbidities, client life circumstances, and treatment type than on mode of delivery. It is important to note that most of the therapies delivered were based on cognitive behavioral therapy or prolonged exposure therapy	• "There were no significant differences in the rates of dropout between the in-person condition and the (home-based) telehealth condition." [25]

Barriers

Multiple barriers were also identified (Table 3), including (1) issues with technology related to connectivity, inconsistent access to a secure, high-quality internet connection, and hardware that disrupts and limits the high quality and secure service delivery [25,52,53,55,64,79,81,85,89], (2) client openness to digital health services [59], (3) challenges to client privacy and comfort, including lack of a quiet, private space, experiences of isolation or disruption in the home environment, and client discomfort with communication over video conferencing [56,59,79,84,85], (4) limits to the therapeutic alliance and therapist comfort with intervention activities that may impact clinical utility and effectiveness [57,58,65,66,79], (5) the ease of abrupt disengagement from treatment and engagement in social avoidant behaviors [55-57,59,79], and (6) safety concerns and risk management [52,59,81].

Table 3. Summary of barriers from all studies.

Themes and sub-themes	Findings	Illustrative quotes
2. Barriers		
2.1 Technological issues are prevalent, causing disruption and limits to high-quality and secure service delivery		
	Connectivity. Challenges because of problematic client and/or therapist internet connection, video conferencing hardware and software, or problems with server connection commonly present difficulties establishing and maintaining a clear, audible, and uninterrupted video-feed impacts the quality of service delivery and client satisfaction	• "The majority of the technical problems that were reported involved lost wireless signals or video or audio quality issues, such as a delay in picture or sound due to poor Internet connection." [25] • "...technical issues with initiating and maintaining a videoconferencing connection were more frequent than expected..." [88]
	Hardware that is compatible for securely connecting with encrypted video conferencing software is not always available for clients. Additionally, as many participants in the studies were provided hardware, more knowledge regarding the protocols and optimal infrastructure for secure delivery of digital health services using personal and/or private computers or video conferencing compatible devices is needed	• "An ideal capability would be to use a network infrastructure that meets U.S. Department of Defense network security requirements but that also allows for the use of privately owned end-user equipment (ie, personal computers, Webcams, mobile devices, etc.)." [88]
2.2 Perceptions of digital health services may limit client acceptance and openness		
	Openness to digital health use may depend on previous experiences or recommendations from trusted individuals or sources. Veterans were described as being hesitant to try new technologies because of issues of security or inconsistency with lifestyle (especially in rural populations). As the studies included clients who were seeking services and open to digital health delivery, more knowledge is yet needed of this population's perceptions and acceptance of digital health services	• "Clinical experience, however, suggests that many patients are hesitant to try new technologies." [55]
2.3 Challenges to client privacy, comfort, and safety exist because of client environment and remote nature of service delivery		
	Lack of a quiet, private space in which clients can engage in therapy without the fear of being overheard by family members or roommates is common	• "... advantages [of digital health] must be balanced by potential shortfalls, such as lack of privacy from family members when televideo sessions are conducted into homes where soundproofing between rooms may not be in place." [87]
	Session disruptions by doorbells or experiencing an abrupt transition back into everyday life after logging off a session made it difficult to engage from the home environment	• "That's why it was hard to switch from talking all about it and then sort of, the hour's up and then you've got to try and get on with normal life." [85]
	Some clients indicated discomfort with communication over video despite satisfaction with their therapist. Concerns about managing strong emotions evoked in therapy in an isolated home environment lead clients to prefer in-person treatment. Additionally, clients may be less trusting of the privacy and confidentiality of digital health services	• "I do not like not knowing who else is in the room with the therapist." [55]
	Safety is difficult to manage in a clinically unsupervised environment where a client may be at risk of purposefully terminating a teleconference session while being at risk of suicide. Much of the reviewed literature excluded clients who posed a risk for suicide, and therefore more examples and knowledge on managing risk and responding to a crisis are necessary. Establishing safety protocols involving family members or neighbors and adjusting service delivery schedules to accommodate is a commonly reported measure; feasibility and ethics in doing so must be considered	• "Potential drawbacks include... the difficulties of ensuring patient safety in a clinically unsupervised environment." [55]
2.4 Limits to the therapeutic alliance and intervention activities may impact clinical utility and effectiveness		
	Establishing and building the therapeutic alliance necessary for effective treatment may be challenged because of the impersonal feeling of videoconferencing, which is influenced by an inability to read all the client and therapist nonverbal body cues	• "Despite being able to see the therapists face, several participants reported that they felt that doing therapy over Skype felt impersonal because they weren't in the same room." [85]

Themes and sub-themes	Findings	Illustrative quotes
	Therapist comfort with digital health may impact the selection of treatment modalities. Further, some clients may benefit from the in-person presence of a clinician to complete exposure activities as per a treatment protocol. Clients with hypervigilance may be unwilling to close eyes during imaginal exposure as they are not reassured that a therapist can watch out for and respond to threats in their environment. Secure exchange of information online related to intake, assessment, and client homework remains an issue	• "Patients who present with more severe symptoms or extreme hypervigilance may be harder to treat via telehealth." [58]
2.5 Ease of disengagement with services and enablement of social avoidant behaviors may be enhanced		
	Clients can disengage quickly and easily if a session becomes too challenging or uncomfortable. They may engage in distractions during the session, such as watching television or browsing the internet	• "if you're having a bad session, you can just switch him off and walk out the room easily." [51]
	Enablement of socially avoidant behaviors may occur when delivering mental health service to a client in their home. Care is required to ensure digital health delivery is not discouraging clients from engaging in healthy life events	• "...Veterans may require leaving their home and attending face-to-face sessions as part of the therapeutic process." [59]

Recommendations

A number of recommendations were identified to support the use of digital health to deliver evidence-based psychotherapy to military members, veterans, and PSP. Key recommendations included (1) identify and manage technological issues that may impede the use of digital health [25,88], (2) supplement interventions to increase patient comfort [75,87], (3) consider ways to establish and maintain rapport and trust [77,89], (4) be flexible and provide additional support as needed to facilitate progress and commitment to therapy [71,77,85], (5) review previously established standards and practices of delivering certain psychotherapeutic interventions to improve suitability for digital delivery [71,87], (6) address risk and safety issues [88], (7) understand and accommodate demographic factors that can influence the client experience of clients using digital health [62,90], and finally (8) support therapists through training to promote their effective use and uptake of digital health [80,84]. Specific findings related to these key themes are presented in Table 4.

Table 4. Recommendations from all studies.

Themes and sub-themes	Findings	Illustrative quotes
3. Recommendations		
3.1 Technological issues that may impede the use of digital health solutions need to be identified and addressed		
	Backups for information technology disruptions. Service providers need alternatives in place if connectivity issues arise that cannot be resolved through technical assistance from a clinician or technical expert	• "If the audio quality remained poor, then the therapist and participant muted their webcams and spoke to each other through the telephone while still using the video feature." [25]
	Secure assessments. Secure methods of distributing and collecting assessments and homework assignments need to be considered	• "Several modifications were also required for sharing homework and study handouts... such as use of screenshots of homework and handouts and holding handouts up to the camera." [88]
3.2 Providers can supplement interventions with pretreatment strategies or peer support to increase patients' comfort with receiving psychotherapy using digital health platforms		
	Pretreatment strategies can help with preparation for therapy and support participants' use of digital health	• "... the present study incorporated many of the recommendations from Gros and colleagues' 2013 review, including the preparation session with a walkthrough and testing of the technology, possibly improving the likelihood of acceptance of and satisfaction with telehealth as a result." [75]
	Peer assistance can support veterans in becoming more open to digital health and play a role in accomplishing the more difficult aspects of treatment	• "... patients who have concerns related to safety or hesitate due to technical concerns may benefit from receiving assistance from a peer before deciding whether or not to try [home telemental health] HTmental health." [59] • "*peer navigators*... may be useful in helping patients to accomplish difficult aspects of treatment, such as in vivo exposure assignments." [71]
3.3 Providers need to consider ways in which rapport and trust can be established and maintained between therapists and clients when using digital health		
	Initial in-person meetings may help to facilitate rapport building for services delivered by digital health	• "... simple changes may result in increased adherence to [prolonged exposure therapy] including... meeting the therapist in-person to increase connection and commitment to the treatment provider." [77]
	Rapport building. Providers should continue to be mindful about embedding ongoing opportunities within therapy to promote rapport building	• "... attention to the development and maintenance of mutually trusting relationships and continued assessment for comfort is recommended." [89]
3.4 Participants may require additional support and flexibility to support progress and commitment to therapy		
	Flexible treatment delivery options and additional information before and during therapy, along with practical solutions to support engagement in digital health appointments, can help with progress and commitment to therapy	• "... offering a hybrid, in-person + telemedicine option may be useful, and would empower patients to match the modality of treatment delivery to the stage of treatment they are completing." [71] • "... participants reporting that they would have liked to have known more before starting the therapy to better prepare... Many participants said the workbooks and additional information given to them between sessions were beneficial." [85] • "... simple changes may result in increased adherence to [prolonged exposure therapy], including using smart-phone calendar reminders; using personal rather than VA-issued smart-phones..." [77]
3.5 Previously established standards and practices of delivering certain psychotherapeutic interventions need to be analyzed to improve suitability for digital health		

Themes and sub-themes	Findings	Illustrative quotes
	Pace of treatment. Due to the independent nature of psychotherapeutic interventions delivered via digital health, there may be a need to alter the pace of treatment	• "... increased hyper-vigilance symptoms in telemedicine vs. in-person PTSD treatment groups, may suggest a need for clinical and administrative modifications to the standard exposure therapy protocol when delivered via telemedicine." [71]
	Variable intervention delivery. To reduce the likelihood of dropping out of therapy because of temporary symptom exacerbation from exposure exercises, intervention delivery may need to be adjusted (massing sessions at the beginning until benefits are experienced) or adjunct with additional cognitive restructuring exercises or education	• "... treatment-interfering cognitions, such as negative treatment expectancies, may need to be addressed with cognitive restructuring in the early stages of treatment." [57]
	Addressing avoidance. In patients where digital health may be reinforcing avoidance behaviors, additional education, and discussion to address avoidance behaviors may be warranted. Peer support and encouragement during in vivo exposure exercises may help to reduce the dropout rate	• "Veterans engaged in [Clinical Video Technology] CVT-delivered PE or [Cognitive Processing Therapy] may need additional education about the role of avoidance in symptom maintenance and frank discussion about. How the CVT modality may be reinforcing avoidance." [57]
3.6 Providers need to consider risk and safety of clients because of the remote and independent nature of digital health		
	Safety planning. Using workable safety standards and planning (including baseline risk assessment, ongoing monitoring of the level of risk, obtaining contact information regarding client's choice of emergency contact before treatment) can facilitate the safe delivery of mental health care to clients in their homes	• "All participants completed a release of information form so that a contact person, of their choice, could assist in case of clinical emergency. The requirements and processes for engaging with third parties were disclosed and discussed during the informed consent process." [88]
3.7 There is a need to recognize and build an understanding of demographic factors (eg, race, gender, and age) that could influence the experience of clients using digital health		
	Demographic considerations. Recognizing the differences in race and ethnic background for clients should be a priority. Gender, age, and related roles may impact an individual's preference and capacity to receive interventions via digital health; considerations around this should be discussed and reviewed with clients	• "Maintaining an understanding of racial obstacles and facilitators in seeking support and continuing follow-up care will be increasingly essential as the military population continues to experience post-traumatic stress related to combat experiences." [62] • "... the effects of age on modality preference among women may reflect barriers to in-office care that uniquely affect the middle-aged and older female populations (eg, responsibility for caring for both older and younger family members, which may make it more difficult to attend office based appointments)." [90] • "Younger women may be more likely to have young children in the home, which may require active caregiving during treatment sessions...Therefore, [office-based treatment] may offer a neutral setting where younger women can receive more private care with fewer distractions." [90]
3.8 Therapists need to be supported through training to promote effective use and uptake of digital health		
	Train providers. It is important to support the training of more therapists across a variety of different settings to use digital health to meet the diverse needs of this population	• "Given the high amount of turnover and transition among providers within and between deployments, it is imperative that all providers using [clinical videoconferencing] technology be briefed prior to, or at the beginning of, deployment." [80] • "... having multiple providers who can offer [Video To Home] decreases burden of trying to meet diverse needs with only one or two designated providers." [63]

Discussion

Efficacy of Digital Health Delivery of Trauma Therapies

This review sought to explore, synthesize, and evaluate the available peer-reviewed research regarding needs, gaps, barriers, and facilitators for safe digital health delivery of trauma-focused psychotherapy for military members, veterans, and/or PSP who had a primary diagnosis of PTSD or a stress-related mental health disorder. The review is important and timely given the current COVID-19 pandemic, the concomitant need for physical distancing, and the ongoing need to support those in rural and remote locations, including theaters of operations or war. There

is an urgent need to ensure that digital mental health services are accessible, effective, safe, and secure.

Overall, there was encouraging support for the use of digital health to facilitate the treatment of trauma-related mental health problems among military members and veterans. The majority of studies included in this review indicated that digital health was as effective as in-person delivery of psychotherapeutic interventions (CPT, PE, and BA-TE) at clinically and statistically significantly reducing PTSD and depressive symptomatology among military members and veterans with mental health challenges. digital health engagement was not rated by clients as inferior to in-person therapy when it came to building a relationship with their clinician [81]. Interestingly, within the included studies, baseline variables such as age, gender, education level, employment status, and relationship status did not appear to correlate with the effectiveness of the psychotherapeutic intervention or the modality of which it was delivered [63,82,84]. Although results regarding adherence to therapy and dropout varied throughout the studies, the modality of therapy delivery (digital health or in-person) did not appear to be a predictor of adherence or dropout.

Importantly, published work confirms that digital health can improve access to treatment in this population because of a combination of reasons including convenience, cost savings, reduced stigma, and comfort and safety of the home environment. We also note that some studies may have included a selection bias by including participants who were already accepting of the idea of remote delivery of services. In the end, personalization and flexibility of therapy, with attention to multiple overlapping facilitators and barriers, may be required.

Quality of the Evidence

The quality of the evidence of the studies included in this review was variable and required a review of multiple quantitative, qualitative, and mixed methods methodologies that made data comparisons challenging. The quality of the RCTs included in this study had PEDro scores ranging from *fair* to *excellent*, with the majority of the studies rated as *good* quality. Increased sample sizes within the RCTs would contribute to larger effect sizes and stronger conclusions regarding the efficacy of digital health psychotherapeutic interventions for this population. Although CPT, PE, and BA-TE were rated as level 1a for evidence, there was a multitude of different outcome measures utilized across all studies, which would make it challenging to further complete meta-analyses with conclusive comparisons. The levels of evidence for the utilization of digital health CBT scored *conflicting*, and we noted a variation in results and a lack of consensus. This finding is similar to other published evidence-based literature focusing on generalized trauma-affected populations [91]. The CASP qualitative checklist purposefully does not result in a score for those assessed studies; however, this tool demonstrated that the qualitative studies were of sound methodology and added a valuable contribution to the literature. Finally, the literature base is relatively new, and as such, it is still fundamentally limited with studies predominantly focusing on 3 therapeutic modalities: PE, CPT, and CBT interventions. Other efficacious psychotherapeutic interventions would benefit from studies with military members, veterans, and/or PSP.

Facilitators, Barriers, and Recommendations

Facilitators, barriers, and recommendations reported in the literature (Tables 2, 3 and 4, respectively) offer clinical and research community insights into ways to respond to the urgent need to research and remotely deliver mental health services. In particular, identified facilitators need to be maximized, barriers overcome, and recommendations implemented for digital health to be best used in practice; doing so will require policy, practice, and system change. As digital health delivery can result in more timely help-seeking behaviors by military members, veterans, and/or PSP, it would be beneficial for more mental health clinicians and mental health clinics to incorporate digital health delivery options as one of several modes of service delivery routinely available to clients [29,31]. This may involve complementing in-person with digital service delivery and varying hours of access to services beyond traditional clinical hours of operation. Doing so would require mental health clinicians and clients to have appropriate technologies, have access to systems and supports, and have strategies and policies in place to ensure connectivity, privacy, security, and confidentiality [84].

Providing immediate support to military members, veterans, and PSP through digital means, particularly when they are contending with operational stresses and trauma, may reduce the acute and long-term impacts on military members, veterans, and PSP themselves and on their teams, organizations, and families [81,89]. Digitally connecting mental health clinicians and remotely or rurally located clients can result in military members, veterans, and PSP receiving support and services that they might not receive otherwise [29,31,55].

The benefits of timely and responsive interventions can reach beyond the individual and into family, organization, and community life [89]. This can positively impact the well-being, operational readiness, and pandemic response. It can also potentially better serve those who might not otherwise seek treatment such as disadvantaged or minority groups, or those who have experienced sexual trauma or operational stress injuries, for example [55,57,59,62,81].

Although the literature regarding the use of digital health to deliver trauma therapies is encouraging, some specific concerns regarding remote delivery warrant consideration. Receiving therapy in one's home, for example, may increase the likelihood of military members, veterans, and PSP avoiding traumatic cues and/or engaging in avoidance behaviors, although this may be more of a dialectic than an actual impediment. That being said, starting with digital health may enable those who would otherwise find it difficult to engage and face their anxieties to actually do so. This may be analogous to the difference between flooding and gradual, systematic desensitization, where digital health may represent the first step for some clients toward challenging, more difficult avoidance behaviors outside the home environment. Similar to the risk management strategies required for in-person service delivery with military members, veterans, and PSP who are at increased risk of self-harm, violence, and suicidality, safeguards need to be implemented

when using digital health–delivered psychotherapeutic interventions. It is important to note that the reviewed literature does not indicate a heightened risk of suicide per se associated with digital health interventions. Consequently, while safety planning is as important for digital health as it is for in-person therapy, suicide risk should not automatically preclude the digital delivery of mental health interventions [76]. Finally, not all patients benefit equally from the digital health interventions. Whereas engaging in therapy at home can be a facilitator for certain clients, some may experience traumatic cues, interruptions, be concerned about privacy, or have discomfort in revealing their home environment to the therapist [77]. This speaks to the importance of determining the right therapy and mode of delivery for the right person at the right time.

Although this scoping review was specific to digital health interventions specific to military member, veterans, and PSP, it may also assist clinicians and researchers in the use of digital health in the general civilian population. It will also add to the growing body of literature exploring digital health utilization for other conditions (ie, substance use, anxiety, depression, dual-diagnosis, schizophrenia, cancer, chronic pain).

Knowledge Gaps and Future Research

With digital health interventions being a relatively recent necessity as a result of the COVID-19 pandemic, this review highlights key gaps in the peer-reviewed literature and areas requiring future research. With regard to user factors, there is a lack of research (ie, only 3 identified studies) related specifically to PSP, emphasizing the need for further study of digital health use in this population. In addition, the majority of studies were conducted in the United States, pointing to the need for future studies to be conducted with military members, veterans, and/or PSP from other international jurisdictions. Further work is also needed not only regarding the impact of demographic and contextual factors but also on the client's stage of treatment, level of functioning, illness severity, attachment style, and comorbidities. How common comorbidities such as depression, personality disorders, or dissociation impact engagement or outcomes in digital health is also yet unknown. For example, some patients require more coregulation in session with the therapist, and there is little information about how digital health delivery might impact this process.

Factors specific to mental health clinicians also warrant further study. This includes consideration of their attitudes, technology acceptance, and usability of digital health. Moreover, little is known about the clinicians' concerns about the impact of digital health on safety, risk management, and control of the therapeutic situation. It is currently unclear whether a clinician's comfort and belief in the efficacy of digital health interventions has a similar influence on therapeutic outcome. It is also important to understand the training and support needs of clinicians to adapt previous in-person interventions to digital health delivery formats. There are also practical aspects to consider regarding the delivery of therapy via digital health. For example, it is currently not well known how the prolonged use of videoconferencing technologies affects a clinician's level of fatigue, attention, and level of engagement throughout a session.

Therapeutic factors associated with the use of digital health also warrant further investigation. This includes factors associated with establishing and maintaining the therapeutic relationship between the clinician and client to those around safety and containment. Similarly, types of treatments that lend themselves better to digital delivery need to be determined and studied. Some therapeutic modalities, for example, are more dependent on in-person therapeutic relational factors (eg, the ability to read body language), whereas others involve experiential activities that may prove difficult to duplicate in a remote delivery context, particularly in the absence of secure platforms to facilitate homework and other barriers. CBT and digital health delivery for military members, veterans, and PSP also warrant additional high-quality research to address the current conflict in the literature. Furthermore, the use of digital health for the delivery of group-based interventions also warrants further study, specifically as it relates to group dynamics, privacy, and stigma, and particularly given the benefits and potential cost-efficiency of group interventions. Future research into individualization of digital health therapy, with attention to multiple facilitators and barriers, is required. Finally, at the macrolevel, further research is needed to determine how to best address systemic issues such as health system policies and medicolegal issues that may present as barriers to digital health for health care organizations, administrators, managers, mental health clinicians, and, therefore, the clients they serve.

Strengths and Limitations

Significant efforts were made to ensure the rigor and quality of the review. Notably, it was performed following a priori and detailed procedures with increased attention to ensure quality control and reduce bias [41,42]. Furthermore, the search strategy was extensive and included 5 databases. Inclusion and exclusion criteria were determined before study onset and adhered to throughout the study. Appropriate calibration and pilot testing, use of at least two independent reviewers for all stages of the process, additional verification of extracted data, and group discussion of conflicts improved the quality of the review.

The authors also acknowledge several limitations of this review. First, the review specifically selected articles focusing on military members, veterans, and/or PSP who had a primary diagnosis of PTSD or a stress-related mental health disorder. Thus, there may be additional informative literature that focuses on other populations. Second, with an imposed date limit, it is possible that quality studies published before 2010 may have been missed; however, with the rapid speed of advancement in digital health, technology, access to technology, and its quality, it is unlikely that research on technology from more than a decade ago before the widespread use of smartphones holds exceptional relevance to this review. Finally, there are limits to aggregate data.

Conclusions

In light of COVID-19, research regarding the remote delivery of trauma therapies for military members, veterans, and PSP is no longer simply novel, but a real-world necessity. In addition, during times of physical distancing, digital health is a mode of delivering psychotherapeutic interventions for those who are in remote locations, including rural areas or during tactical and

operational deployment. Despite some promising evidence in currently published literature, health care organizations and mental health clinicians should continue to proceed cautiously with remote delivery of psychotherapeutic trauma therapies as much research is still needed to address *the digital divide* among trauma-affected military members, veterans, and/or PSP. This systematic scoping review included 38 studies researching factors related to the use of digital health to deliver psychotherapeutic interventions for military members, veterans, and/or PSP affected by trauma. Evidence for the effectiveness of digital delivery of PE, CPT, and BA-TE on military members, veterans, and PSP was rated level 1a, whereas evidence for CBT was conflicting. The narrative synthesis suggests that digital health delivery of these therapies can be as effective as in-person delivery; for some, it may increase access, reduce stigma, and facilitate engagement with mental health clinicians. Issues of risk, safety, privacy and security, digital health modality, clinician factors, and barriers to digital health need to be researched further as does the potential for additional trauma therapies, including ART, ADT, and EMDR, to be suitable for digital delivery. Further elucidation is also needed for gender, racial, and cultural factors that may create differences in client outcomes, preferences, and needs. Professional organizations are likely needed to invest in training for clinicians, formulate guidelines for digital health delivery of trauma therapies, formulate ethical and medicolegal guidance, and assist in adapting to the digital health environment.

Acknowledgments

The authors would like to thank the Canadian Institute of Health Research for funding support provided through the Knowledge Synthesis: COVID-19 in the Mental Health & Substance Use grant competition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy.
[DOCX File , 17 KB - mhealth_v8i9e22079_app1.docx]

Multimedia Appendix 2
Detailed descriptive analysis of studies included in the review.
[DOCX File , 1162 KB - mhealth_v8i9e22079_app2.docx]

Multimedia Appendix 3
Descriptive analysis and outcomes of all studies included in the scoping review.
[DOCX File , 36 KB - mhealth_v8i9e22079_app3.docx]

References

1. Torous J, Keshavan M. COVID-19, mobile health and serious mental illness. Schizophr Res 2020 Apr;218:36-37 [FREE Full text] [doi: 10.1016/j.schres.2020.04.013] [Medline: 32327314]
2. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
3. Adalja AA, Toner E, Inglesby TV. Priorities for the US health community responding to COVID-19. J Am Med Assoc 2020 Mar 3;323(14):1343-1344 epub ahead of print. [doi: 10.1001/jama.2020.3413] [Medline: 32125355]
4. Glossary of Terms. CIPSRT – Canadian Institute for Public Safety Research and Treatment. 2019. URL: https://www.cipsrt-icrtsp.ca/glossary-of-terms/ [accessed 2020-06-06]
5. Ricciardelli R, Czarnuch S, Carleton RN, Gacek J, Shewmake J. Canadian public safety personnel and occupational stressors: how PSP interpret stressors on duty. Int J Environ Res Public Health 2020 Jul 1;17(13):4736 [FREE Full text] [doi: 10.3390/ijerph17134736] [Medline: 32630259]
6. McFarlane AC. The impact of war on mental health: lest we forget. World Psychiatry 2015 Oct;14(3):351-353. [doi: 10.1002/wps.20253] [Medline: 26407792]
7. Fischer H. A Guide to US Military Casualty Statistics: Operation Freedoms Sentinel, Operation Inherent Resolve, Operation New Dawn, Operation Iraqi Freedom, and Operation Enduring Freedom. Congressional Research Service. 2015. URL: www.crs.gov [accessed 2020-07-05]
8. Derluyn I, Broekaert E, Schuyten G, de Temmerman E. Post-traumatic stress in former Ugandan child soldiers. Lancet 2004 Mar 13;363(9412):861-863. [doi: 10.1016/S0140-6736(04)15734-6] [Medline: 15031032]

9. Vermetten E, Greenberg N, Boeschoten MA, Delahaije R, Jetly R, Castro CA, et al. Deployment-related mental health support: comparative analysis of NATO and allied ISAF partners. Eur J Psychotraumatol 2014;5:- [FREE Full text] [doi: 10.3402/ejpt.v5.23732] [Medline: 25206953]
10. Veterans Affairs Canada: Joint Suicide Prevention Strategy. Government of Canada. 2017. URL: https://www.canada.ca/content/dam/dnd-mdn/documents/reports/2017/caf-vac-joint-suicide-prevention-strategy.pdf [accessed 2020-07-06]
11. Strong Secure Engaged: Canada's Defense Policy. Government of Canada. 2017. URL: https://www.canada.ca/en/department-national-defence/corporate/policies-standards/canada-defence-policy.html [accessed 2020-06-06]
12. American Psychological Association. Depressive disorders. In: Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Publishing; 2013.
13. Hoencamp R, Vermetten E, Tan EC, Putter H, Leenen LP, Hamming JF. Systematic review of the prevalence and characteristics of battle casualties from NATO coalition forces in Iraq and Afghanistan. Injury 2014 Jul;45(7):1028-1034. [doi: 10.1016/j.injury.2014.02.012] [Medline: 24878294]
14. Jinkerson JD. Defining and assessing moral injury: a syndrome perspective. Traumatology 2016 Jun;22(2):122-130. [doi: 10.1037/trm0000069]
15. Koenig HG, Youssef NA, Ames D, Teng EJ, Hill TD. Examining the overlap between moral injury and PTSD in US veterans and active duty military. J Nerv Ment Dis 2020 Jan;208(1):7-12. [doi: 10.1097/NMD.0000000000001077] [Medline: 31738222]
16. Milliken CS, Auchterlonie JL, Hoge CW. Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war. J Am Med Assoc 2007 Nov 14;298(18):2141-2148. [doi: 10.1001/jama.298.18.2141] [Medline: 18000197]
17. Wagner A, Jakupcak M. Combat-related stress reactions among US veterans of wartime service. In: The Oxford Handbook of Military Psychology. Oxford, UK: Oxford University Press; 2012.
18. Zamorski MA, Boulos D. The impact of the military mission in Afghanistan on mental health in the Canadian armed forces: a summary of research findings. Eur J Psychotraumatol 2014;5:- [FREE Full text] [doi: 10.3402/ejpt.v5.23822] [Medline: 25206951]
19. Jones N, Seddon R, Fear NT, McAllister P, Wessely S, Greenberg N. Leadership, cohesion, morale, and the mental health of UK armed forces in Afghanistan. Psychiatry 2012;75(1):49-59. [doi: 10.1521/psyc.2012.75.1.49] [Medline: 22397541]
20. Eekhout I, Reijnen A, Vermetten E, Geuze E. Post-traumatic stress symptoms 5 years after military deployment to Afghanistan: an observational cohort study. Lancet Psychiatry 2016 Jan;3(1):58-64. [doi: 10.1016/S2215-0366(15)00368-5] [Medline: 26681368]
21. Australian Guidelines for the Treatment of Acute Stress Disorder & Posttraumatic Stress Disorder. Phoenix Australia. 2013. URL: https://phoenixaustralia.org/wp-content/uploads/2015/03/%0APhoenix-ASD-PTSD-Guidelines.pdf [accessed 2020-07-06]
22. Fulton JJ, Calhoun PS, Wagner HR, Schry AR, Hair LP, Feeling N, et al. The prevalence of posttraumatic stress disorder in operation enduring freedom/operation Iraqi freedom (OEF/OIF) veterans: a meta-analysis. J Anxiety Disord 2015 Apr;31:98-107. [doi: 10.1016/j.janxdis.2015.02.003] [Medline: 25768399]
23. van Til L, Sweet J, Poirier A. Well-being of Canadian regular force veterans : findings from LASS 2016 survey. Government of Canada Publications. 2017. URL: http://publications.gc.ca/pub?id=9.839366&sl=0 [accessed 2020-06-06]
24. Carleton RN, Afifi TO, Turner S, Taillieu T, Duranceau S, LeBouthillier DM, et al. Mental disorder symptoms among public safety personnel in Canada. Can J Psychiatry 2018 Jan;63(1):54-64 [FREE Full text] [doi: 10.1177/0706743717723825] [Medline: 28845686]
25. Yuen EK, Gros DF, Price M, Zeigler S, Tuerk PW, Foa EB, et al. Randomized controlled trial of home-based telehealth versus in-person prolonged exposure for combat-related PTSD in veterans: preliminary results. J Clin Psychol 2015 Jun;71(6):500-512. [doi: 10.1002/jclp.22168] [Medline: 25809565]
26. Eftekhari A, Ruzek JI, Crowley JJ, Rosen CS, Greenbaum MA, Karlin BE. Effectiveness of national implementation of prolonged exposure therapy in veterans affairs care. JAMA Psychiatry 2013 Sep;70(9):949-955. [doi: 10.1001/jamapsychiatry.2013.36] [Medline: 23863892]
27. Resick P, Monson CG. Psychosocial Treatments for PTSD. Guilford Press. 2007. URL: http://www.guilford.com [accessed 2020-07-06]
28. Stecker T, McHugo G, Xie H, Whyman K, Jones M. RCT of a brief phone-based CBT intervention to improve PTSD treatment utilization by returning service members. Psychiatr Serv 2014 Oct;65(10):1232-1237 [FREE Full text] [doi: 10.1176/appi.ps.201300433] [Medline: 24933496]
29. van Voorhees BW, Gollan J, Fogel J. Pilot study of Internet-based early intervention for combat-related mental distress. J Rehabil Res Dev 2012;49(8):1175-1190 [FREE Full text] [doi: 10.1682/jrrd.2011.05.0095] [Medline: 23341310]
30. Cully JA, Jameson JP, Phillips LL, Kunik ME, Fortney JC. Use of psychotherapy by rural and urban veterans. J Rural Health 2010;26(3):225-233. [doi: 10.1111/j.1748-0361.2010.00294.x] [Medline: 20633090]
31. van Voorhees BW, Ellis JM, Gollan JK, Bell CC, Stuart SS, Fogel J, et al. Development and process evaluation of a primary care internet-based intervention to prevent depression in emerging adults. Prim Care Companion J Clin Psychiatry 2007;9(5):346-355 [FREE Full text] [doi: 10.4088/pcc.v09n0503] [Medline: 17998953]

32. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
33. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
34. Kitchiner NJ, Lewis C, Roberts NP, Bisson JI. Active duty and ex-serving military personnel with post-traumatic stress disorder treated with psychological therapies: systematic review and meta-analysis. Eur J Psychotraumatol 2019;10(1):1684226 [FREE Full text] [doi: 10.1080/20008198.2019.1684226] [Medline: 31762951]
35. Lewis C, Roberts NP, Andrew M, Starling E, Bisson JI. Psychological therapies for post-traumatic stress disorder in adults: systematic review and meta-analysis. Eur J Psychotraumatol 2020;11(1):1729633 [FREE Full text] [doi: 10.1080/20008198.2020.1729633] [Medline: 32284821]
36. Steenkamp MM, Litz BT, Marmar CR. First-line psychotherapies for military-related PTSD. J Am Med Assoc 2020 Jan 30:- epub ahead of print. [doi: 10.1001/jama.2019.20825] [Medline: 31999301]
37. VA/DOD Clinical Practice Guideline for the Management of Posttraumatic Stress Disorder and Acute Stress Disorder: Clinician Summary. Focus - Psychiatry Online. 2018. URL: https://focus.psychiatryonline.org/doi/10.1176/appi.focus.16408 [accessed 2020-07-06]
38. Clinical Practice Guideline For the Treatment of PTSD. American Psychological Association (APA). 2017. URL: https://www.apa.org/ptsd-guideline/ptsd.pdf [accessed 2020-08-20]
39. Gray MJ, Schorr Y, Nash W, Lebowitz L, Amidon A, Lansing A, et al. Adaptive disclosure: an open trial of a novel exposure-based intervention for service members with combat-related psychological stress injuries. Behav Ther 2012 Jun;43(2):407-415. [doi: 10.1016/j.beth.2011.09.001] [Medline: 22440075]
40. Kip KE, Rosenzweig L, Hernandez DF, Shuman A, Sullivan KL, Long CJ, et al. Randomized controlled trial of accelerated resolution therapy (ART) for symptoms of combat-related post-traumatic stress disorder (PTSD). Mil Med 2013 Dec;178(12):1298-1309. [doi: 10.7205/MILMED-D-13-00298] [Medline: 24306011]
41. Neubauer NA, Lapierre N, Ríos-Rincón A, Miguel-Cruz A, Rousseau J, Liu L. What do we know about technologies for dementia-related wandering? A scoping review: examen de la portée : que savons-nous à propos des technologies de gestion de l'errance liée à la démence? Can J Occup Ther 2018 Jun;85(3):196-208. [doi: 10.1177/0008417418777530] [Medline: 29972049]
42. Cruz AM, Rincón AM, Dueñas WR, Torres DA, Bohórquez-Heredia AF. What does the literature say about using robots on children with disabilities? Disabil Rehabil Assist Technol 2017 Jul;12(5):429-440. [doi: 10.1080/17483107.2017.1318308] [Medline: 28440095]
43. Recent Software Updates. EndNote: Clarivate Analytics. URL: https://endnote.com/downloads/available-updates/ [accessed 2020-07-06]
44. Covidence Systematic Review Software. Covidence - Better Systematic Review Management. URL: https://www.covidence.org/reviews/active [accessed 2020-07-06]
45. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
46. Popay J, Roberts H, Sowden A. Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme. ESRC Methods Program. 2006. URL: https://www.researchgate.net/publication/233866356_Guidance_on_the_conduct_of_narrative_synthesis_in_systematic_reviews_A_product_from_the_ESRC_Methods_Programme#fullTextFileContent [accessed 2020-07-06]
47. Pope C, Mays N, Popay J. Synthesizing Qualitative and Quantitative Health Evidence: A Guide to Methods. New York, USA: Open University Press; 2007.
48. Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the physiotherapy evidence database (PEDro). Aust J Physiother 2002;48(1):43-49 [FREE Full text] [doi: 10.1016/s0004-9514(14)60281-6] [Medline: 11869164]
49. Critical Appraisal Skills Programme: CASP. 2018. URL: https://casp-uk.net/ [accessed 2020-07-06]
50. Aubut J, Teasell R, Cullen N, Marshall SB. Evidence-Based Review of Moderate to Severe Brain Injury. Semantic Scholar: AI-Powered Research Tool. 2013. URL: https://pdfs.semanticscholar.org/2b08/beaff788b361f6e3f5ca37ed8fda616b3fa9.pdf [accessed 2020-08-20]
51. Straus S, Richardson W, Glasziou P, Haynes R. Evidence-Based Medicine: How to Practice and Teach EBM. Third Edition. Edinburgh, UK: SAGE Publications; 2005.
52. Ziemba SJ, Bradley NS, Landry LP, Roth CH, Porter LS, Cuyler RN. Posttraumatic stress disorder treatment for operation enduring freedom/operation Iraqi freedom combat veterans through a civilian community-based telemedicine network. Telemed J E Health 2014 May;20(5):446-450. [doi: 10.1089/tmj.2013.0312] [Medline: 24617961]
53. Wierwille JL, Pukay-Martin ND, Chard KM, Klump MC. Effectiveness of PTSD telehealth treatment in a VA clinical sample. Psychol Serv 2016 Nov;13(4):373-379. [doi: 10.1037/ser0000106] [Medline: 27657798]
54. Whealin J, Yoneda A, Nelson D, Hilmes T, Kawasaki M, Yan O. A culturally adapted family intervention for rural Pacific Island veterans with PTSD. Psychol Serv 2017 Aug;14(3):295-306. [doi: 10.1037/ser0000186] [Medline: 28805414]

55. Whealin JM, King L, Shore P, Spira J. Diverse veterans' pre- and post-intervention perceptions of home telemental health for posttraumatic stress disorder delivered via tablet. Int J Psychiatry Med 2017 Jan;52(1):3-20. [doi: 10.1177/0091217417703291] [Medline: 28486881]
56. Wells SY, Glassman LH, Talkovsky AM, Chatfield MA, Sohn MJ, Morland LA, et al. Examining changes in sexual functioning after cognitive processing therapy in a sample of women trauma survivors. Womens Health Issues 2019;29(1):72-79. [doi: 10.1016/j.whi.2018.10.003] [Medline: 30455090]
57. Valentine LM, Donofry SD, Broman RB, Smith ER, Rauch SA, Sexton MB. Comparing PTSD treatment retention among survivors of military sexual trauma utilizing clinical video technology and in-person approaches. J Telemed Telecare 2020;26(7-8):443-451. [doi: 10.1177/1357633X19832419] [Medline: 30975048]
58. Tuerk PW, Yoder M, Ruggiero KJ, Gros DF, Acierno R. A pilot study of prolonged exposure therapy for posttraumatic stress disorder delivered via telehealth technology. J Trauma Stress 2010 Feb;23(1):116-123 [FREE Full text] [doi: 10.1002/jts.20494] [Medline: 20135675]
59. Trahan LH, Carges E, Stanley MA, Evans-Hudnall G. Decreasing PTSD and depression symptom barriers to weight loss using an integrated CBT approach. Clin Case Stud 2016 Mar 17;15(4):280-294. [doi: 10.1177/1534650116637918]
60. Strachan M, Gros DF, Ruggiero KJ, Lejuez CW, Acierno R. An integrated approach to delivering exposure-based treatment for symptoms of PTSD and depression in OIF/OEF veterans: preliminary findings. Behav Ther 2012 Sep;43(3):560-569. [doi: 10.1016/j.beth.2011.03.003] [Medline: 22697444]
61. Stecker T, Shiner B, Watts BV, Jones M, Conner KR. Treatment-seeking barriers for veterans of the Iraq and Afghanistan conflicts who screen positive for PTSD. Psychiatr Serv 2013 Mar 1;64(3):280-283. [doi: 10.1176/appi.ps.001372012] [Medline: 23450385]
62. Stecker T, Adams L, Carpenter-Song E, Nicholson J, Streltzov N, Xie H. Intervention efficacy in engaging Black and White veterans with post-traumatic stress disorder into treatment. Soc Work Public Health 2016 Oct;31(6):481-489 [FREE Full text] [doi: 10.1080/19371918.2016.1160340] [Medline: 27210830]
63. Morland LA, Mackintosh M, Rosen CS, Willis E, Resick P, Chard K, et al. Telemedicine versus in-person delivery of cognitive processing therapy for women with posttraumatic stress disorder: a randomized noninferiority trial. Depress Anxiety 2015 Nov;32(11):811-820. [doi: 10.1002/da.22397] [Medline: 26243685]
64. Morland LA, Mackintosh M, Greene CJ, Rosen CS, Chard KM, Resick P, et al. Cognitive processing therapy for posttraumatic stress disorder delivered to rural veterans via telemental health: a randomized noninferiority clinical trial. J Clin Psychiatry 2014 May;75(5):470-476. [doi: 10.4088/JCP.13m08842] [Medline: 24922484]
65. Morland LA, Mackintosh M, Glassman LH, Wells SY, Thorp SR, Rauch SA, et al. Home-based delivery of variable length prolonged exposure therapy: a comparison of clinical efficacy between service modalities. Depress Anxiety 2020 Apr;37(4):346-355. [doi: 10.1002/da.22979] [Medline: 31872563]
66. Morland LA, Hynes AK, Mackintosh M, Resick PA, Chard KM. Group cognitive processing therapy delivered to veterans via telehealth: a pilot cohort. J Trauma Stress 2011 Aug;24(4):465-469. [doi: 10.1002/jts.20661] [Medline: 21793047]
67. Maieritsch KP, Smith TL, Hessinger JD, Ahearn EP, Eickhoff JC, Zhao Q. Randomized controlled equivalence trial comparing videoconference and in person delivery of cognitive processing therapy for PTSD. J Telemed Telecare 2016 Jun;22(4):238-243. [doi: 10.1177/1357633X15596109] [Medline: 26231819]
68. Luxton DD, Pruitt LD, O'Brien K, Kramer G. An evaluation of the feasibility and safety of a home-based telemental health treatment for posttraumatic stress in the US military. Telemed J E Health 2015 Nov;21(11):880-886. [doi: 10.1089/tmj.2014.0235] [Medline: 26103565]
69. Franklin CF, Walton JL, Raines AM, Chambliss JL, Corrigan SA, Cuccurullo LJ, et al. Pilot study comparing telephone to in-person delivery of cognitive-behavioural therapy for trauma-related insomnia for rural veterans. J Telemed Telecare 2018 Oct;24(9):629-635. [doi: 10.1177/1357633X17732366] [Medline: 28950755]
70. 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]
71. Hernandez-Tejada MA, Zoller JS, Ruggiero KJ, Kazley AS, Acierno R. Early treatment withdrawal from evidence-based psychotherapy for PTSD: telemedicine and in-person parameters. Int J Psychiatry Med 2014;48(1):33-55. [doi: 10.2190/PM.48.1.d] [Medline: 25354925]
72. Hernandez-Tejada MA, Hamski S, Sánchez-Carracedo D. Incorporating peer support during in vivo exposure to reverse dropout from prolonged exposure therapy for posttraumatic stress disorder: clinical outcomes. Int J Psychiatry Med 2017;52(4-6):366-380. [doi: 10.1177/0091217417738938] [Medline: 29179664]
73. Grubbs KM, Fortney JC, Pyne JM, Hudson T, Moore WM, Custer P, et al. Predictors of initiation and engagement of cognitive processing therapy among veterans with PTSD enrolled in collaborative care. J Trauma Stress 2015 Dec;28(6):580-584. [doi: 10.1002/jts.22049] [Medline: 26625355]
74. Gros DF, Yoder M, Tuerk PW, Lozano BE, Acierno R. Exposure therapy for PTSD delivered to veterans via telehealth: predictors of treatment completion and outcome and comparison to treatment delivered in person. Behav Ther 2011 Jun;42(2):276-283. [doi: 10.1016/j.beth.2010.07.005] [Medline: 21496512]

75. Gros DF, Lancaster CL, López CM, Acierno R. Treatment satisfaction of home-based telehealth versus in-person delivery of prolonged exposure for combat-related PTSD in veterans. J Telemed Telecare 2018 Jan;24(1):51-55. [doi: 10.1177/1357633X16671096] [Medline: 27672059]
76. Gallegos AM, Streltzov NA, Stecker T. Improving treatment engagement for returning operation enduring freedom and operation iraqi freedom veterans with posttraumatic stress disorder, depression, and suicidal ideation. J Nerv Ment Dis 2016 May;204(5):339-343 [FREE Full text] [doi: 10.1097/NMD.0000000000000489] [Medline: 26894313]
77. Franklin CL, Cuccurullo L, Walton JL, Arseneau JR, Petersen NJ. Face to face but not in the same place: a pilot study of prolonged exposure therapy. J Trauma Dissociation 2017;18(1):116-130. [doi: 10.1080/15299732.2016.1205704] [Medline: 27348462]
78. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. JAMA Psychiatry 2015 Jan;72(1):58-67. [doi: 10.1001/jamapsychiatry.2014.1575] [Medline: 25409287]
79. Seal KH, Abadjian L, McCamish N, Shi Y, Tarasovsky G, Weingardt K. A randomized controlled trial of telephone motivational interviewing to enhance mental health treatment engagement in Iraq and Afghanistan veterans. Gen Hosp Psychiatry 2012;34(5):450-459. [doi: 10.1016/j.genhosppsych.2012.04.007] [Medline: 22632925]
80. Pelton D, Wangelin B, Tuerk P. Utilizing telehealth to support treatment of acute stress disorder in a theater of war: prolonged exposure via clinical videoconferencing. Telemed J E Health 2015 May;21(5):382-387. [doi: 10.1089/tmj.2014.0111] [Medline: 25764266]
81. Olden M, Wyka K, Cukor J, Peskin M, Altemus M, Lee FS, et al. Pilot study of a telehealth-delivered medication-augmented exposure therapy protocol for PTSD. J Nerv Ment Dis 2017 Feb;205(2):154-160 [FREE Full text] [doi: 10.1097/NMD.0000000000000563] [Medline: 27441461]
82. Murphy D, Turgoose D. Evaluating an internet-based video cognitive processing therapy intervention for veterans with PTSD: a pilot study. J Telemed Telecare 2019 Jun 17:1357633X1985039. [doi: 10.1177/1357633x19850393]
83. Morland L, Wells S, Glassman L, Grubbs K, Mackintosh M, Golshan S, et al. What do veterans want? Understanding veterans' preferences for PTSD treatment delivery. Mil Med 2019 Dec 1;184(11-12):686-692. [doi: 10.1093/milmed/usz035] [Medline: 30839067]
84. Boykin DM, Keegan F, Thompson KE, Voelkel E, Lindsay JA, Fletcher TL. Video to home delivery of evidence-based psychotherapy to veterans with posttraumatic stress disorder. Front Psychiatry 2019;10:893 [FREE Full text] [doi: 10.3389/fpsyt.2019.00893] [Medline: 31920747]
85. Ashwick R, Turgoose D, Murphy D. Exploring the acceptability of delivering cognitive processing therapy (CPT) to UK veterans with PTSD over Skype: a qualitative study. Eur J Psychotraumatol 2019;10(1):1573128 [FREE Full text] [doi: 10.1080/20008198.2019.1573128] [Medline: 30774784]
86. Acierno R, Knapp R, Tuerk P, Gilmore AK, Lejuez C, Ruggiero K, et al. A non-inferiority trial of prolonged exposure for posttraumatic stress disorder: in person versus home-based telehealth. Behav Res Ther 2017 Feb;89:57-65 [FREE Full text] [doi: 10.1016/j.brat.2016.11.009] [Medline: 27894058]
87. Acierno R, Gros DF, Ruggiero KJ, Hernandez-Tejada MA, Knapp RG, Lejuez CW, et al. Behavioral activation and therapeutic exposure for posttraumatic stress disorder: a noninferiority trial of treatment delivered in person versus home-based telehealth. Depress Anxiety 2016 May;33(5):415-423. [doi: 10.1002/da.22476] [Medline: 26864655]
88. Luxton DD, Pruitt LD, O'Brien K, Kramer G. An evaluation of the feasibility and safety of a home-based telemental health treatment for posttraumatic stress in the US military. Telemed J E Health 2015 Nov;21(11):880-886. [doi: 10.1089/tmj.2014.0235] [Medline: 26103565]
89. Whealin J, Yoneda A, Nelson D, Hilmes T, Kawasaki M, Yan O. A culturally adapted family intervention for rural Pacific Island veterans with PTSD. Psychol Serv 2017 Aug;14(3):295-306. [doi: 10.1037/ser0000186] [Medline: 28805414]
90. Morland L, Wells S, Glassman L, Grubbs K, Mackintosh M, Golshan S, et al. What do veterans want? Understanding veterans' preferences for PTSD treatment delivery. Mil Med 2019 Dec 1;184(11-12):686-692. [doi: 10.1093/milmed/usz035] [Medline: 30839067]
91. Internet-Delivered Cognitive Behavioural Therapy for Post-Traumatic Stress Disorder: Recommendations. Canadian Agency for Drugs and Technologies in Health (CADTH). 2020. URL: https://cadth.ca/sites/default/files/hta-he/op0537-icbt-for-ptsd-recommendations.pdf [accessed 2020-07-04]

Abbreviations

ADT: adaptive disclosure therapy
ART: accelerated resolution therapy
BA: behavioral activation
BA-TE: behavioral activation and therapeutic exposure
CASP: Critical Skills Appraisal Program
CBT: cognitive behavioral therapy
CPT: cognitive processing therapy

EMDR: eye movement desensitization reprocessing
PE: prolonged exposure therapy
PEDro: Physiotherapy Evidence Database
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSP: public safety personnel
PTSD: posttraumatic stress disorder
PTSI: posttraumatic stress injury
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 07.07.20; peer-reviewed by B Smith, J Salisbury; comments to author 22.07.20; revised version received 28.07.20; accepted 01.08.20; published 21.09.20.

Please cite as:
Jones C, Miguel-Cruz A, Smith-MacDonald L, Cruikshank E, Baghoori D, Kaur Chohan A, Laidlaw A, White A, Cao B, Agyapong V, Burback L, Winkler O, Sevigny PR, Dennett L, Ferguson-Pell M, Greenshaw A, Brémault-Phillips S
Virtual Trauma-Focused Therapy for Military Members, Veterans, and Public Safety Personnel With Posttraumatic Stress Injury: Systematic Scoping Review
JMIR Mhealth Uhealth 2020;8(9):e22079
URL: http://mhealth.jmir.org/2020/9/e22079/
doi:10.2196/22079
PMID:32955456

Original Paper

Clinometric Gait Analysis Using Smart Insoles in Patients With Hemiplegia After Stroke: Pilot Study

Minseok Seo[1], MA, MD; Myung-Jun Shin[2], MD, PhD; Tae Sung Park[3], MA; Jong-Hwan Park[3], PhD

[1]School of Medicine, Pusan National University, Busan, Republic of Korea

[2]Department of Rehabilitation Medicine, School of Medicine, Pusan National University, Busan, Republic of Korea

[3]Biomedical Research Institute, Pusan National University Hospital, Busan, Republic of Korea

Corresponding Author:
Myung-Jun Shin, MD, PhD
Department of Rehabilitation Medicine
School of Medicine
Pusan National University
305 Gudeok-Ro Seo-Gu
Busan, 49241
Republic of Korea
Phone: 82 1085130907
Email: drshinmj@gmail.com

Abstract

Background: For effective rehabilitation after stroke, it is essential to conduct an objective assessment of the patient's functional status. Several stroke severity scales have been used for this purpose, but such scales have various limitations.

Objective: Gait analysis using smart insole technology can be applied continuously, objectively, and quantitatively, thereby overcoming the shortcomings of other assessment tools.

Methods: To confirm the reliability of gait analysis using smart insole technology, normal healthy controls wore insoles in their shoes during the Timed Up and Go (TUG) test. The gait parameters were compared with the manually collected data. To determine the gait characteristics of patients with hemiplegia due to stroke, they were asked to wear insoles and take the TUG test; gait parameters were calculated and compared with those of control subjects. To investigate whether the gait analysis accurately reflected the patients' clinical condition, we analyzed the relationships of 22 gait parameters on 4 stroke severity scales.

Results: The smart insole gait parameter data were similar to those calculated manually. Among the 18 gait parameters tested, 14 were significantly effective at distinguishing patients from healthy controls. The smart insole data revealed that the stance duration on both sides was longer in patients than controls, which has proven difficult to show using other methods. Furthermore, the sound side in patients showed a markedly longer stance duration. Regarding swing duration, that of the sound side was shorter in patients than controls, whereas that of the hemiplegic side was longer. We identified 10 significantly correlated gait parameters on the stroke severity scales. Notably, the difference in stance duration between the sound and hemiplegic sides was significantly correlated with the Fugl-Meyer Assessment (FMA) lower extremity score.

Conclusions: This study confirmed the feasibility and applicability of the smart insole as a device to assess the gait of patients with hemiplegia due to stroke. In addition, we demonstrated that the FMA score was significantly correlated with the smart insole data. Providing an environment where stroke patients can easily measure walking ability helps to maintain chronic functions as well as acute rehabilitation.

Trial Registration: UMIN Clinical Trials Registry UMIN000041646, https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000047538

(JMIR Mhealth Uhealth 2020;8(9):e22208) doi:10.2196/22208

KEYWORDS

stroke; hemiplegia; gait; smart insole; medical informatics; rehabilitation; observational; wearable; assessment

Introduction

Stroke remains one of the leading causes of disease burden worldwide [1]. Despite efforts to prevent stroke and reduce its impact with early intervention, many people live with persistent, chronic deficits that require significant rehabilitation [2]. The challenge for stroke rehabilitation is to decrease impairments and promote patient activity and participation by optimizing early outcome prognosis and therapeutic care [3].

To ensure effective rehabilitation therapy for a patient diagnosed with stroke, it is essential to perform an objective assessment of the patient's functional status. For this purpose, various functional tools have been developed for evaluating patients diagnosed with stroke, such as the Fugl-Meyer Assessment (FMA), the Mini–Mental State Examination (MMSE) for body functions, the Modified Barthel Index (MBI) for activities, and the Stroke Impact Scale for participation, which are the most widely used according to the International Classification of Functioning, Disability, and Health (ICF). However, these scales are time-consuming to administer and problematic due to the influence of the subjective perception of the evaluator [4]. The ordinal scale of these instruments represents a further limitation. Furthermore, it is difficult to record data continuously while administering treatments.

Instrumental gait analysis can be used to evaluate patients continuously, objectively, and quantitatively [5]. Moreover, it can overcome the limitations of other scales, as evaluation and treatment can be simultaneous. We investigated whether in-depth gait analysis using smart insole technology reflects the functional status of patients with hemiplegia due to stroke.

Methods

Experimental Design

To investigate the reliability of gait analysis using the smart insole device, healthy control subjects wore the insoles in their shoes and completed a Timed Up and Go (TUG) test. The spatiotemporal gait parameters were compared with the values calculated by manually measuring the TUG time and step count. To determine the gait characteristics of patients with hemiplegia due to stroke, they were asked to wear the insoles and take the TUG. The gait parameters were calculated and compared with those of control subjects. The TUG differentiates subjects with chronic stroke from healthy elderly subjects and shows test-retest reliability [6]. Therefore, this study employed the TUG test for gait analysis of patients diagnosed with stroke. To examine whether the gait analysis data accurately reflected patients' clinical conditions, we analyzed the relationship between gait parameters and the stroke severity scale data.

Participants

Participants in this study included 10 healthy control subjects and 10 patients with hemiplegia due to stroke. Included patients were all of chronic-stage status, were diagnosed with stroke, understood the purpose of the experiment, and could walk independently without the use of a walking aid during the TUG test. Patients were excluded if they had musculoskeletal disorders that could affect gait or diabetic complications that could cause peripheral neuropathy. Table 1 displays the characteristics of the 10 patients included in the study.

Table 1. Basic characteristics of patients (N=10).

Characteristics	Value, mean (range)
Duration of illness (months)	47.0 (12.0-85.0)
Height (cm)	165.3 (148.0-175.0)
Weight (kg)	69.4 (49.6-97.0)
MMSE[a]	24.4 (14.0-30.0)
MBI[b]	59.0 (22.0-89.0)
FMA_Uex[c]	25.5 (4.0-47.0)
FMA_Lex[d]	19.0 (7.0-28.0)

[a]MMSE: Mini–Mental State Examination.

[b]MBI: Modified Barthel Index.

[c]FMA_Uex: Fugl-Meyer Assessment, upper-extremity score.

[d]FMA_Lex: Fugl-Meyer Assessment, lower-extremity score.

Measurements

The following functional tools were used to evaluate patients: (1) the Timed Up and Go (TUG) test, (2) the Modified Barthel Index (MBI), (3) the Fugle-Meyer Assessment (FMA), and (4) the Mini–Mental State Examination (MMSE).

Timed Up and Go (TUG) Test

The test was performed according to a standard protocol on a standard TUG track (Figure 1) [7]. The patients wore the smart insoles in their shoes. Each insole is equipped with 8 pressure sensors, a 3-axis accelerometer, and a gyroscope; the data were measured at a frequency of 100 Hz (Figure 2).

Figure 1. Timed Up and Go (TUG) test.

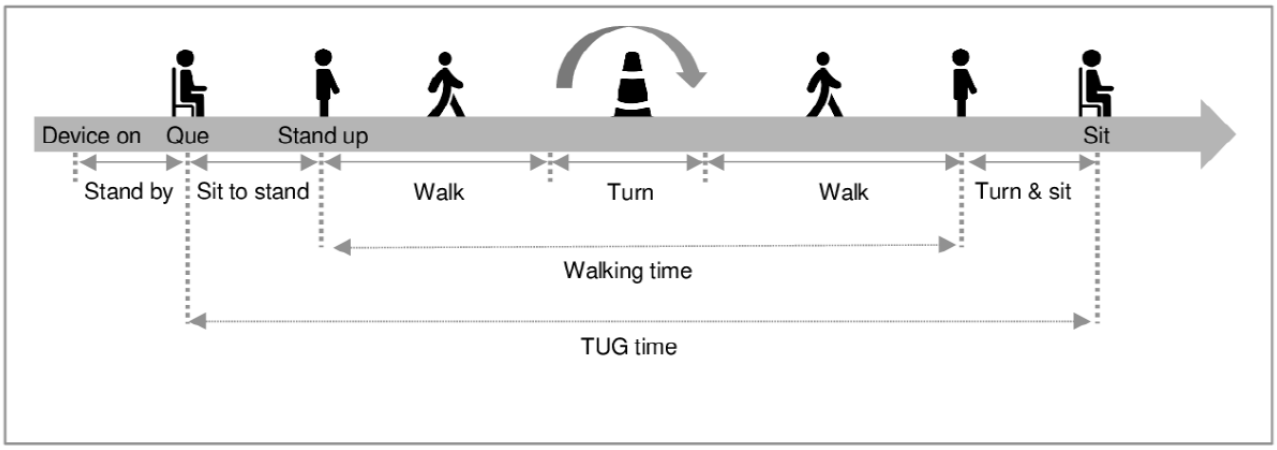
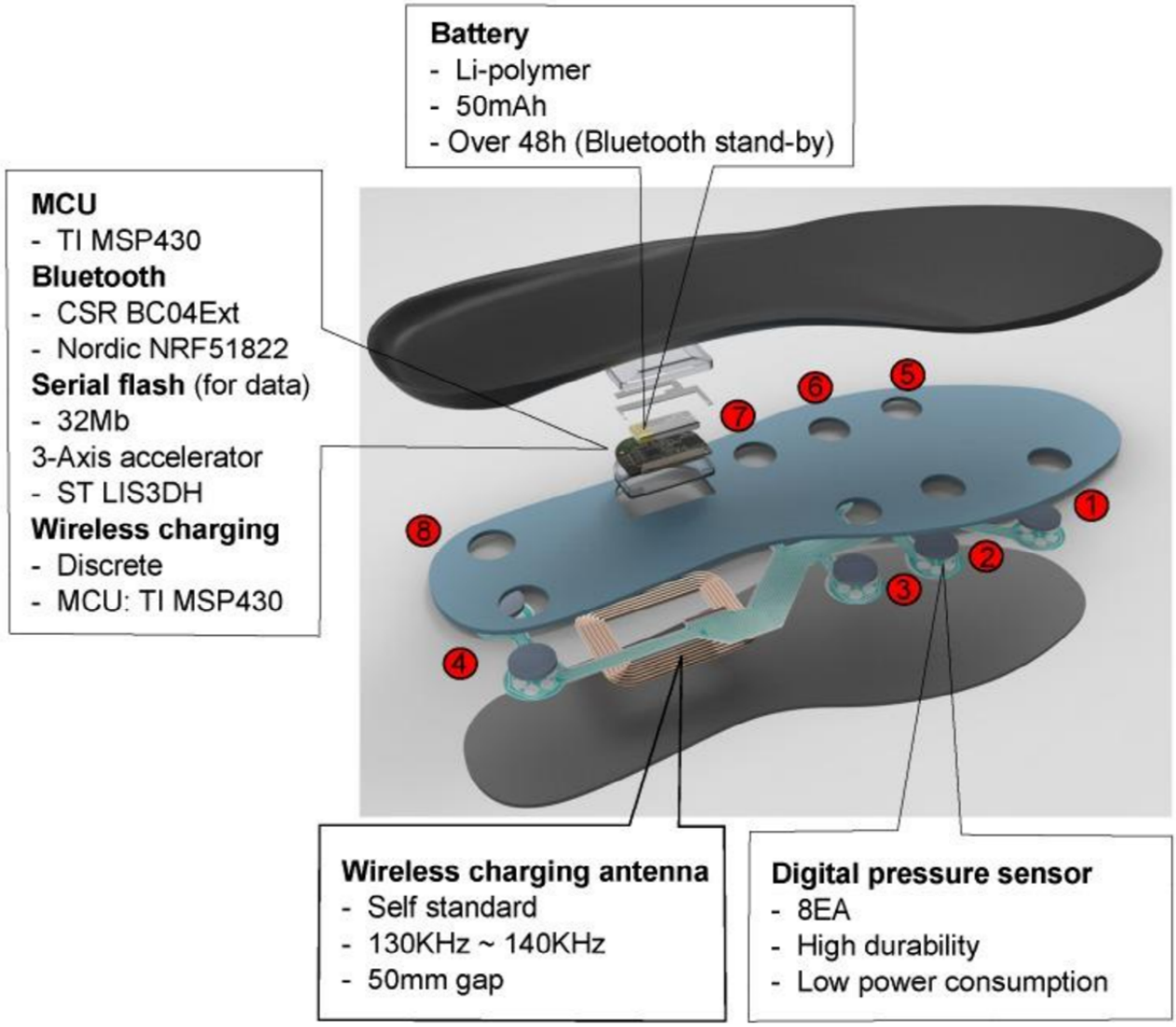


Figure 2. Design of the smart insole sensor module. EA: each; h: hour; kHz: kilohertz; mAh: milliampere hour; Mb: megabytes; MCU: microcontrol unit; mm: millimeter.



Modified Barthel Index (MBI)

The most widely used tool for evaluating daily activities of stroke patients in the current rehabilitation is the MBI and the functional independence measure (FIM). Because the FIM has restrictions (such as a fee to the copyright holder and the need for training), the MBI is more commonly used in clinical practice and research [8]. The MBI also has a metrological advantage, which is reported to be more sensitive, simpler, easier to score than other tools, more reliable, and more feasible

[9,10]. These benefits have contributed to ensuring that the MBI is translated and valid in many countries [8,11-13].

Fugl-Meyer Assessment (FMA)

The FMA is a 226-point scale developed to evaluate recovery from hemiplegic stroke. In stroke rehabilitation, it is one of the most comprehensive quantitative measurements of motor impairment (body function). Although the use of the FMA for patients with mild motor impairment is limited by a ceiling effect, the FMA is reliable and is highly recommended as a body impairment scale based on available evidence [14]. Using the FMA in combination with a general activity measure such as the MBI or the TUG may provide additional information to improve the measurement of recovery for stroke patients.

Mini–Mental State Examination (MMSE)

The MMSE is frequently used in clinical practice. Although this instrument was originally developed to screen for dementia and delirium, the use of the MMSE has been extended, and many studies now use it as a screening instrument for global cognitive impairment [15]. The Montreal Cognitive Assessment is the best candidate to predict recovery; however, the MMSE is still a useful scale in clinical settings of stroke rehabilitation. As cognitive dysfunction affects learning and rehabilitation outcomes, as well as predicting functional independence after stroke, assessment of cognitive function must also be considered to evaluate the severity of stroke [16].

Data Handling and Analyses

During the experiment, the sensor data were stored in the flash memory in the insole and transmitted via Bluetooth after the experiment. After collecting the data, we performed noise filtering and differentiated between the swing and stance phases by reference to the total number of activated pressure sensors. As shown in Figure 3, the swing phase corresponded to when the total number of activated pressure sensors was 0. The stance was represented by non-zero values, as was described previously by Truong et al [17], who used the same equipment and concluded that their experimental tests performed accurately to distinguish between swing and stance phases. We calculated the gait parameters for each participant, as shown in Table 2. The single support time is when only one foot is in the stance phase, and the double support time is when both feet are in the stance phase. The percentages of difference between swing and stance durations were calculated by dividing the differences in swing and stance durations (measured in seconds) by the corresponding gait cycle duration (measured in seconds). A 2-sample *t* test was used to compare the gait parameters between patients with hemiplegia and normal control subjects, and the coefficient of determination (R^2) was used to analyze the Pearson correlation between gait parameters and stroke severity scale results. The calculations were performed using R statistical software [18].

Figure 3. Dividing the swing and stance phase. If the individual press sensor has a pressure of 4.3 N/cm2 or more, it is defined as activated pressure sensor. The swing phase corresponds to when the total number of activated pressure sensors is 0; the stance is represented by non-zero values. If both left and right sides are in stance state, it is treated as double support; if only one side is in stance state, it is treated as single support.

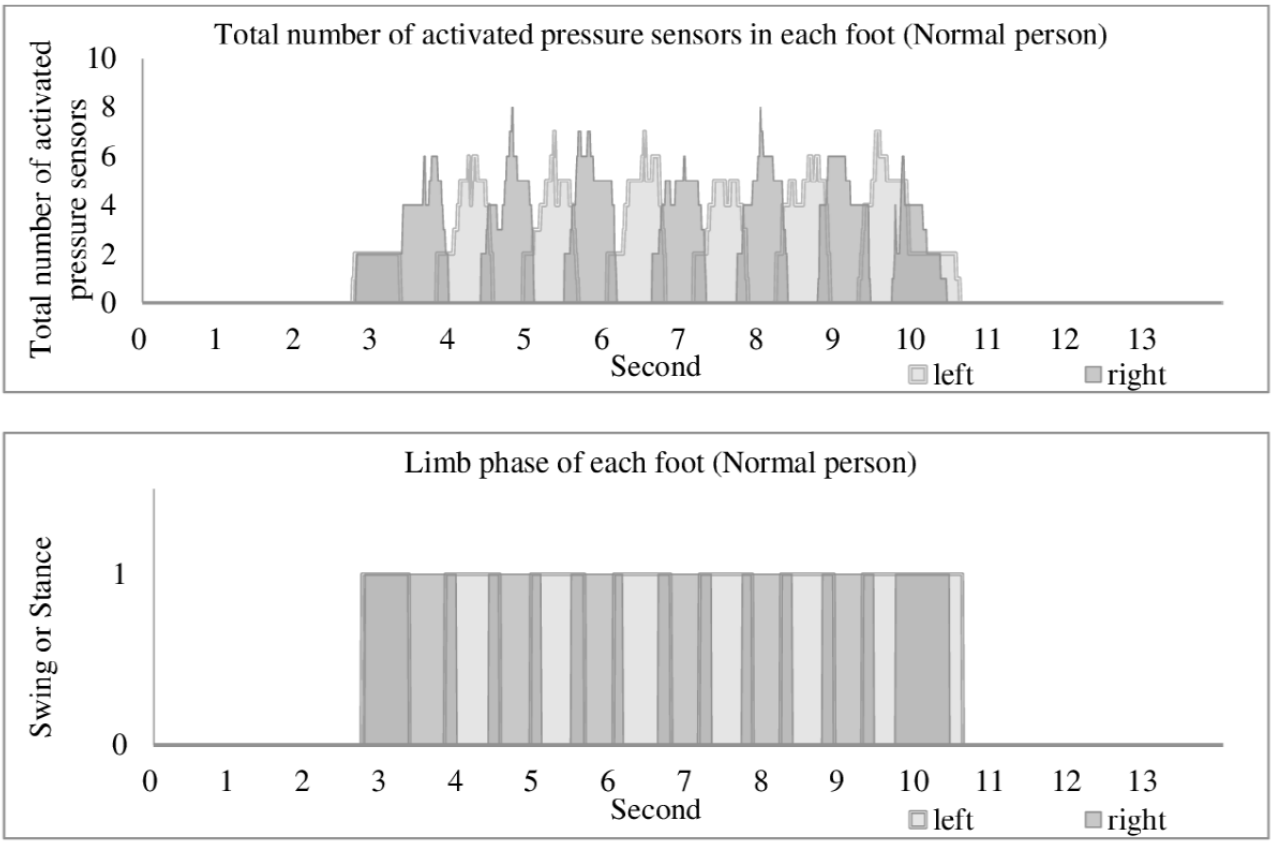


Table 2. Comparison of gait parameters between patients with hemiplegia (n=10) and normal control subjects (n=10), using a 2-sample *t* test.

Gait parameters	Hemiplegia, mean (range)	Normal control, mean (range)	*P* value
Walking speed (meters/second)	0.32 (0.20-0.48)	0.88 (0.73-1.02)	<.001
Stride length (meters)	0.40 (0.23-0.57)	0.91 (0.80-1.00)	<.001
TUG time (seconds)	24.69 (15.92-36.70)	9.83 (8.62-11.26)	<.001
Walking time (seconds)	20.63 (12.48-30.47)	6.95 (5.91-8.24)	<.001
Single support time, SD (%)	0.14 (0.08,0.23)	0.04 (0.02,0.06)	<.001
Single support time, SD (seconds)	0.19 (0.08-0.36)	0.04 (0.03-0.07)	.001
Double support time, mean (seconds)	0.19 (0.13-0.25)	0.13 (0.10-0.16)	.002
Difference in swing duration (%)	0.18 (0.01-0.46)	0.02 (0.00-0.04)	.005
Difference in stance duration (%)	0.19 (0.00-0.45)	0.03 (0.01-0.06)	.005
Difference in stance duration (seconds)	0.27 (0.00-0.56)	0.03 (0.01-0.06)	.005
Difference in swing duration (seconds)	0.25 (0.01-0.54)	0.02 (0.00-0.04)	.006
Double support time, mean (%)	0.15 (0.10-0.21)	0.12 (0.10-0.13)	.03
Sit to standing (seconds)	4.06 (1.19-6.23)	2.88 (2.36-3.62)	.03
Cadence (steps/min)	97.25 (66.64-138.09)	114.49 (97.50-124.60)	.048
Double support time, SD (s)	0.08 (0.02-0.19)	0.05 (0.03-0.05)	.053
Single support time, mean (%)	0.35 (0.30-0.41)	0.38 (0.35-0.41)	.08
Double support time, SD (%)	0.06 (0.02-0.17)	0.04 (0.03-0.05)	.17
Single support time, mean (seconds)	0.47 (0.30-0.66)	0.40 (0.36-0.46)	.19

Ethics Approval and Consent to Participate

The Institutional Review Board of Pusan National University Hospital approved this study, which was registered retrospectively (Pusan National University Hospital, https://www.pnuh.or.kr; 1812-010-074). This study was registered in the UMIN Clinical Trials Registry (University hospital Medical Information Network, https://www.umin.ac.jp/ctr/; No. UMIN000041646).

Results

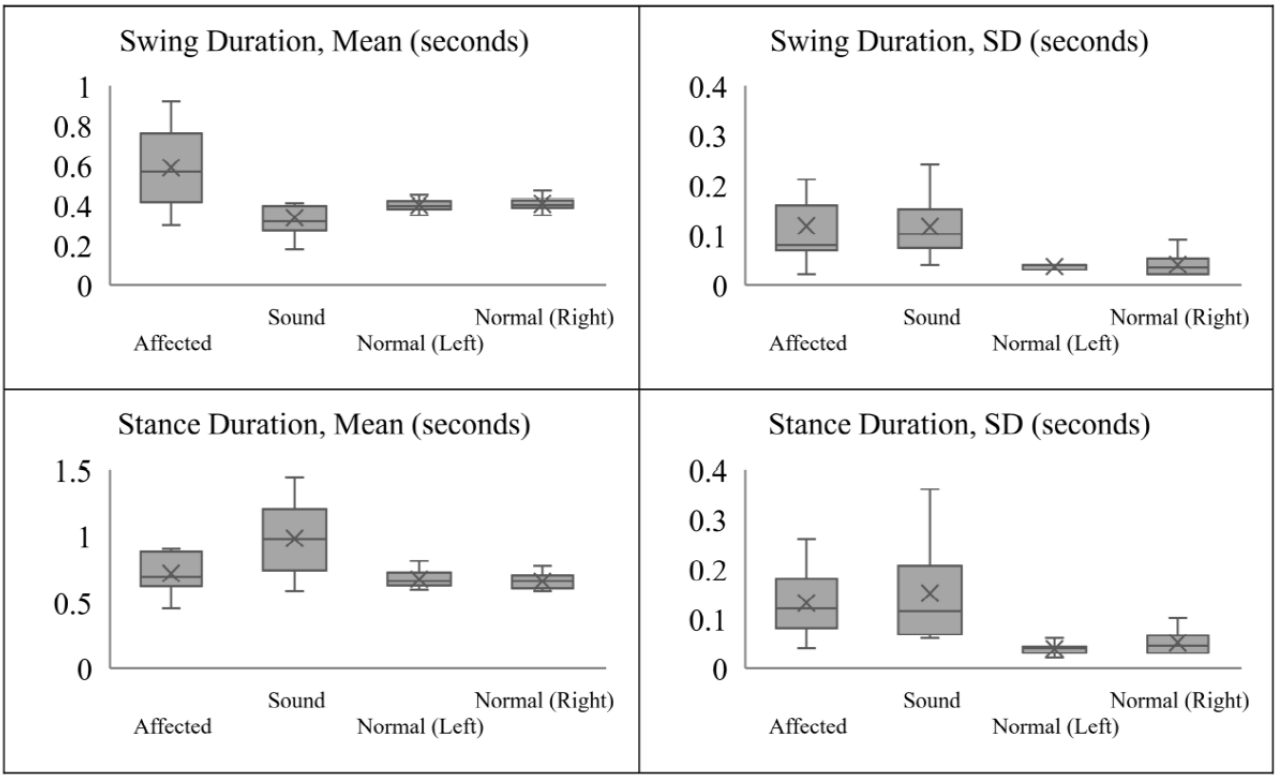
Control Group

The smart insole results for the control subjects were as follows: TUG time = 8.62–11.26 seconds, cadence = 97.5–124.6 steps/min, walking speed = 0.73–1.02 meters/second, and stride length = 0.80–1.00 meters.

Patients Diagnosed With Stroke

The patients diagnosed with stroke had a mean duration of disease of 47.0 (SD 29.4) months and a mean MMSE score of 24.4 (SD 4.9). The gait parameters that showed significant differences between patients with hemiplegia and normal control subjects were the TUG time (seconds), walking time (seconds), sit-to-standing time (seconds), cadence (steps/min), walking speed (meters/second), stride length (meters), standard deviation of single support time (seconds, percentage), mean double support time (seconds, percentage), difference in swing duration between the sound and hemiplegic sides (seconds, percentage) and difference in stance duration between the sound and hemiplegic sides (seconds, percentage) (Table 2). The patients had a mean TUG time of 24.69 seconds, which was longer than that of the control subjects (9.83 seconds). The walking speed range of the patients with hemiplegia was 0.20 to 0.48 meters per second, which was significantly slower than that of control subjects (0.73-1.02 meters per second). The cadence of the patients was 66.64-138.09 steps per minute, which was more broadly distributed than that of the control subjects (97.50-124.60 steps per min), and their stride length was 0.23-0.57 meters, which was significantly shorter than that of control subjects (0.80-1.00 meters). The results showed that the patients had a slower walking speed due to shorter stride length and higher cadence. The average single support time was 0.47 seconds (mean value of the hemiplegic and sound sides), which did not differ significantly from that of control subjects (0.40 seconds). However, the mean difference in stance duration between the sound and hemiplegic sides was 0.27 seconds, which was significantly greater than that of control subjects (0.03 seconds) (Figure 4).

Figure 4. Boxplot displaying the swing and stance duration distribution of patients' hemiplegic and sound sides, based on a 5-number summary: minimum, maximum, median, first quartile, and third quartile. X: mean. The patients showed a shorter swing duration on the sound side than the control subjects, whereas that on the hemiplegic side was longer. Patients showed a longer stance duration on both sides compared to the control subjects; however, the sound side showed a markedly longer duration.



Correlations Between Gait Parameters and Stroke Severity Scale Results

As shown in Table 3 and Figure 5, the difference in stance duration between the sound and hemiplegic sides was most strongly correlated with the FMA lower extremity score in the hemiplegia group ($y = -38.64x + 26.39$, $R^2 = 0.59$). The recalculated R^2 value, excluding an outlier patient, was larger at 0.71, indicating a clear correlation between the FMA lower extremity score and the difference in stance duration between the sound and hemiplegic sides.

The 4 stroke severity scales used for correlation analysis were the FMA for the lower extremity and the upper extremity, the MMSE, and the MBI. The 22 gait parameters used for correlation analysis were the TUG time (seconds), walking time (seconds), sit-to-standing time (seconds), cadence (steps per minute), walking speed (meters per second), stride length (meters), mean single support time (seconds), standard deviation of single support time (seconds), mean double support time (seconds), standard deviation of double support time (seconds), difference in swing duration (seconds), difference in stance duration (seconds), percentage difference in swing duration, percentage difference in stance duration, mean swing duration of hemiplegic side (seconds), standard deviation of swing duration of hemiplegic side (seconds), mean swing duration of sound side (seconds), standard deviation of swing duration of sound side (seconds), mean stance duration of hemiplegic side (seconds), standard deviation of stance duration of hemiplegic side (seconds), mean sound-side stance duration (seconds), and standard deviation of sound-side stance duration (seconds).

Table 3. Correlations between gait parameters and stroke severity scales.

Gait parameter	Stroke severity scale	R^2	*P* value[a]
Difference in stance duration (%)	FMA_Lex[b]	0.591	.009
Difference in stance duration (%)	FMA_Uex[c]	0.539	.02
Difference in swing duration (%)	FMA_Lex[b]	0.517	.02
Sound-side swing duration, SD (seconds)	MMSE[d]	0.498	.02
Cadence (steps/min)	MMSE[d]	0.484	.03
Hemiplegic-side stance duration, SD (seconds)	MMSE[d]	0.452	.03
Single support time, mean (seconds)	MMSE[d]	0.448	.03
Sound-side stance duration, mean (seconds)	MMSE[d]	0.442	.04
Walking speed (meters/second)	MMSE[d]	0.403	.048
Single support time, SD (seconds)	MMSE[d]	0.401	.049

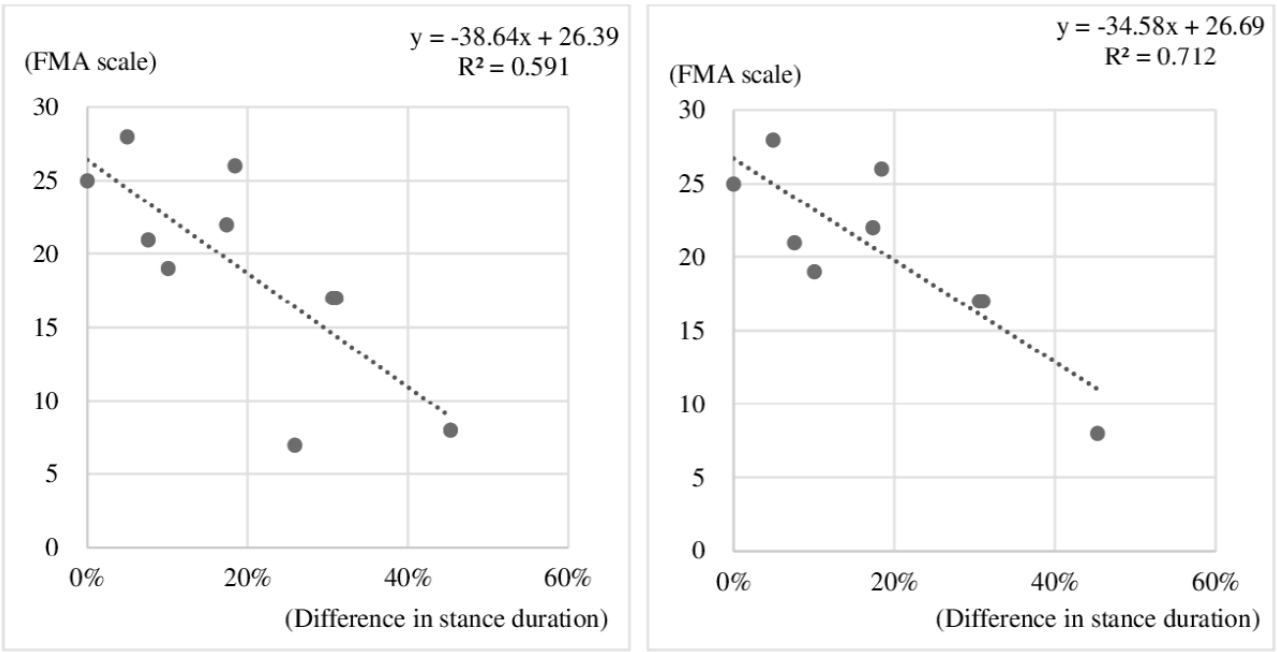
[a]Non-significant results omitted.

[b]FMA_Lex: Fugl-Meyer Assessment, lower-extremity score.

[c]FMA_Uex: Fugl-Meyer Assessment, upper-extremity score.

[d]MMSE: Mini–Mental State Examination.

Figure 5. Correlation between the difference in stance duration and the Fugl-Meyer Assessment (FMA) lower-extremity score. The recalculated R^2 value, excluding an outlier patient, was larger at 0.712.



Discussion

Feasibility and Applicability of the Smart Insole as a Gait Analysis Device

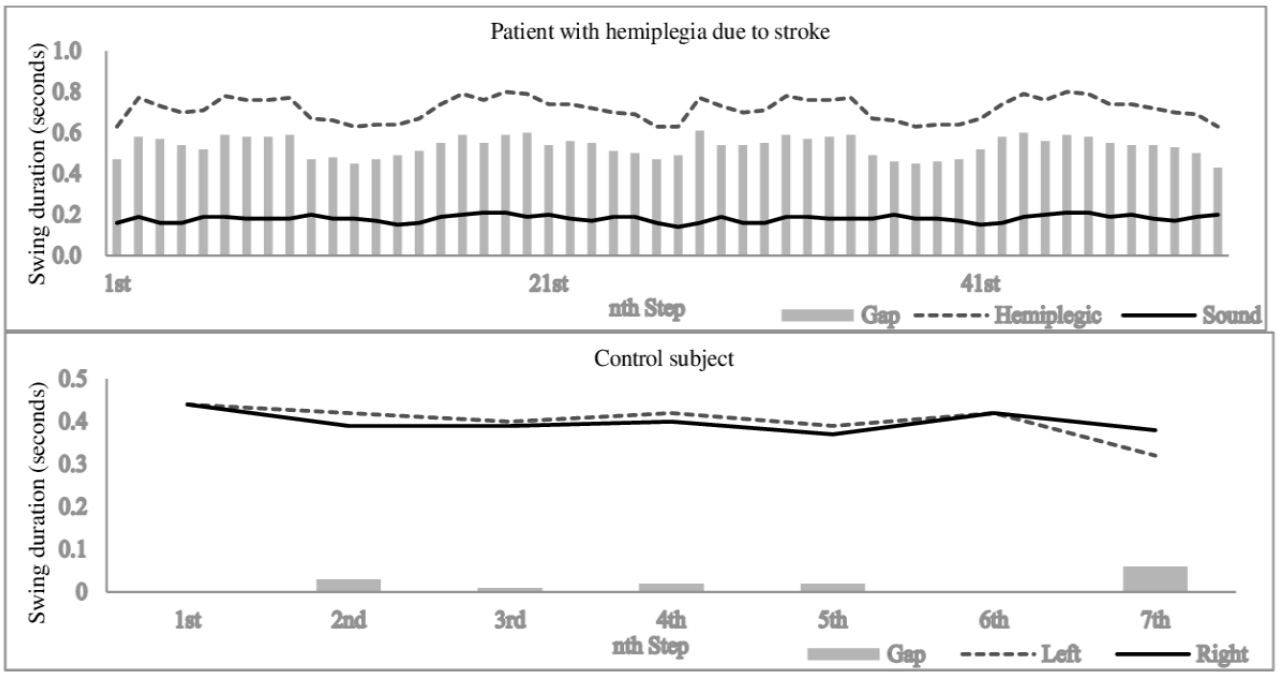
As a gait analysis device, the smart insole showed that it may be used in place of the manual calculations of gait parameters made by a clinician, such as calculations of TUG time and step count. Truong et al [17] concluded that this equipment made accurate estimates of walking distance with a mean walking distance estimation error of 4.8% and 3.1% for 16 meters and 89 meters walking distance, respectively. David et al [19] demonstrated the feasibility and applicability of rehabilitation using an eSHOE system similar to the smart insole device. The smart insole used in this study was also applied to predict energy consumption in a recent article [20]. Thus, we suggest that these types of systems are more useful than other methods as they are location-independent and can measure additional parameters that have not been addressed in the past.

Gait Features of Patients With Hemiplegia due to Stroke

Analysis of the smart insole data revealed several gait characteristics of patients with hemiplegia due to stroke that

were difficult to identify using previous methods. First, the patients showed slower walking speeds and shorter stride lengths than the control subjects. To compensate for these differences, some patients showed a faster cadence than control subjects. As shown in Figure 4, we observed a difference between the hemiplegic and sound sides. While the patients showed a longer stance duration on both sides compared to control subjects, the sound side showed a markedly longer duration. In addition, the patients showed a shorter swing duration on the sound side than the control subjects, whereas that on the hemiplegic side was longer. It is likely that the hemiplegic side moved slowly during the swing phase and lacked the strength required to support the body's weight during the stance phase. Figure 6 shows the difference in swing duration between the sound and hemiplegic sides for a patient and between the left and right sides for a control subject. The control subject showed a difference close to zero (ie, the left- and right-sided swing durations were almost the same). However, the patient showed a longer swing duration on the hemiplegic side than the sound side, with the difference exceeding 0.4 seconds. In an analysis of the relationship between TUG performance and gait parameters, Bonnyaud et al [21] revealed the importance of swing and stance duration, where the motor ability of the paretic lower limb and the single support phase on the paretic side determined TUG performance. The results obtained with the smart insole were consistent with the previous study and could be obtained more easily.

Figure 6. Difference in swing duration between a patient with hemiplegia and a normal control subject. The control subject showed a difference close to zero (ie, the left- and right-sided swing durations were almost the same); however, the patient showed a longer swing duration on the hemiplegic side than the sound side, with the difference exceeding 0.4 seconds.



Gait Analysis for Clinical Assessment

We examined the correlations of 22 gait parameters with stroke severity scale data, such as FMA lower- and upper- extremity scores, and MMSE and MBI scores. Of the 22 gait parameters, 10 showed significant correlations (Table 3). The strongest correlations were found between the FMA lower-extremity score and the difference in stance duration between the sound and hemiplegic sides.

As the MMSE is used for measuring cognitive impairment, it seems that it has an overall correlation with several gait parameters rather than having a high correlation with specific gait parameters. In this study, the MMSE does not have a high correlation with a specific gait parameter like the FMA lower extremity score does. However, the MMSE has a significant correlation with most types (7) of gait parameters (Table 3). We expect that the variability of gait parameters could be related to cognitive impairment due to a decrease in the efficiency of rehabilitation training. In this study, we found that patients diagnosed with stroke have a greater standard deviation for gait parameters, which is more meaningful than the average (Table 2, Figure 4). Thus, if there were patients with the same motor impairment but with more severe cognitive impairments, they would possibly have been distinguished by the variability of gait parameters during TUG, a complex task.

MBI does not have a clear correlation with specific gait parameters because it evaluates the performance of daily life activities, and not just walking ability. In contrast, the FMA is an instrument for measuring recovery from sensorimotor stroke [22], and it can evaluate various body parts and functions. Thus, the FMA should intuitively have a direct connection with walking ability.

Guzika et al [23] emphasized the importance of gait symmetry, for which only parameters measuring gait symmetry were associated with the degree of gait control in patients after stroke diagnosis. In this respect, it is reasonable that differences in stance duration and swing duration between the sound and hemiplegic sides would show the most significant correlations with FMA lower-extremity scores, which reflect gait ability.

In addition, Hiengkaew et al [24] reported that both the FMA lower-extremity score and TUG were reliable measurements of postural balance and lower limb movements in individuals with chronic stroke. As the stability of the swing and stance phases determine TUG performance [21], and the FMA lower-extremity score has a significant correlation with the difference in swing and stance duration, a correlation between FMA lower-extremity score and TUG could be deduced. No direct correlation was found between TUG time and the FMA lower-extremity score in this experiment; however, it was found that gait parameters during TUG had a significant correlation with the FMA lower-extremity score (Table 3). As the TUG includes a return section, it is more complex than the walking test for straight distances and can reflect the problem of hemiplegic impairment well. Thus, despite the small number of subjects in this study, the gait parameters during TUG appear to have a significant correlation with the FMA lower-extremity scores. However, if we apply TUG in a conventional way, time information would have been the only information we could have.

We further investigated the outlier patient (Figure 5), who had chronic hemiplegia for 6 years. Although the patient had a relatively low FMA score, the difference in swing duration between the sound and hemiplegic sides was not significant. While the TUG time and cadence of the patient were average for the patient group, the prolonged mean support time suggested instability. Although there were no specific symptoms or signs at the time of the experiment, the patient had a history of treatments for poststroke seizure, which was the only difference from other patients. However, it was difficult to explain the clear reason. We compared the FMA and gait analysis results of the patient with medical records held by other rehabilitation specialists and noted differences in opinion indicating underestimation of the FMA results relative to function. Despite the small number of subjects, the reason for accounting for the outlier is not to assert a clear correlation but rather to show potential for a new method. Future studies with larger numbers of patients are important for further analysis of the characteristics of such outlier patients.

Future Clinical Applications

Yu et al [25] developed an ensemble model that can estimate the 33-item FMA upper-extremity score based on measured values using wearable devices. Gait is more difficult to analyze as it requires more complex interactions of nerves and muscles; however, it seems to be possible to estimate each item of existing functional evaluations through in-depth gait analysis in follow-up studies.

Gait analysis using the smart insole can be performed without temporal or geographical constraints and might be useful for promoting appropriate gait patterns. This study showed that the smart insole measuring device could replace existing functional evaluations such as the FMA. The smart insole could be used for biofeedback training, both at home and in the hospital environment. Providing an environment where stroke patients can easily measure walking ability may help to maintain chronic functions as well as acute rehabilitation.

Conclusions

We introduce the feasibility and utility of the smart insole for assessing gait features in patients diagnosed with hemiplegia. In addition, we found that the results for the FMA functional index, which is the most commonly used instrument for assessing patients with motor impairment, were significantly correlated with those obtained using the smart insole. Further studies are required to confirm the clinical effectiveness of the smart insole for rehabilitation treatment and long-term monitoring of patients after a stroke diagnosis.

Conflicts of Interest

None declared.

References

1. Feigin VL, Krishnamurthi RV, Parmar P, Norrving B, Mensah GA, Bennett DA, et al. Update on the Global Burden of Ischemic and Hemorrhagic Stroke in 1990-2013: The GBD 2013 Study. Neuroepidemiology 2015 Oct 28;45(3):161-176. [doi: 10.1159/000441085]
2. Benjamin E, Virani S, Callaway C, Chamberlain A, Chang A, Cheng S, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. Circulation 2018 Mar 20;137(12):e67-e492. [doi: 10.1161/CIR.0000000000000558] [Medline: 29386200]
3. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: The Stroke Recovery and Rehabilitation Roundtable taskforce. International Journal of Stroke 2017 Jul 12;12(5):444-450. [doi: 10.1177/1747493017711816]
4. Hamacher D, Singh N, Van Dieën J, Heller M, Taylor W. Kinematic measures for assessing gait stability in elderly individuals: a systematic review. J. R. Soc. Interface 2011 Aug 31;8(65):1682-1698. [doi: 10.1098/rsif.2011.0416]
5. Tamburini P, Mazzoli D, Stagni R. Towards an objective assessment of motor function in sub-acute stroke patients: Relationship between clinical rating scales and instrumental gait stability indexes. Gait & Posture 2018 Jan;59:58-64. [doi: 10.1016/j.gaitpost.2017.09.033]
6. Ng SS, Hui-Chan CW. The Timed Up & Go Test: Its Reliability and Association With Lower-Limb Impairments and Locomotor Capacities in People With Chronic Stroke. Archives of Physical Medicine and Rehabilitation 2005 Aug;86(8):1641-1647. [doi: 10.1016/j.apmr.2005.01.011]

7. Podsiadlo D, Richardson S. The timed. J Am Geriatr Soc 1991 Feb;39(2):142-148. [Medline: 1991946]
8. Jung HY, Park BK, Shin HS, Kang YK, Pyun SB, Paik NJ. Development of the Korean Version of Modified Barthel Index (K-MBI): Multi-center Study for Subjects with Stroke. Journal of the Korean Academy of Rehabilitation Medicine 2007;31(3):283-297 The text of this paper is written in Korean [FREE Full text]
9. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. Journal of Clinical Epidemiology 1989 Jan;42(8):703-709. [doi: 10.1016/0895-4356(89)90065-6]
10. Duffy L, Gajree S, Langhorne P, Stott DJ, Quinn TJ. Reliability (Inter-rater Agreement) of the Barthel Index for Assessment of Stroke Survivors. Stroke 2013 Feb;44(2):462-468. [doi: 10.1161/strokeaha.112.678615]
11. Leung SO, Chan CC, Shah S. Development of a Chinese version of the Modified Barthel Index — validity and reliability. Clin Rehabil 2016 Jul;21(10):912-922. [doi: 10.1177/0269215507077286]
12. Küçükdeveci AA, Yavuzer G, Tennant A, Süldür N, Sonel B, Arasil T. Adaptation of the modified Barthel Index for use in physical medicine and rehabilitation in Turkey. Scand J Rehabil Med 2000 Jun;32(2):87-92 [FREE Full text] [Medline: 10853723]
13. Shah S, Cooper B, Maas F. The Barthel Index and ADL Evaluation in Stroke Rehabilitation in Australia, Japan, the UK and the USA. Aust Occup Ther J 1992 Mar;39(1):5-13. [doi: 10.1111/j.1440-1630.1992.tb01729.x] [Medline: 21790640]
14. 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 2016 Jun 30;16(3):232-240. [doi: 10.1177/154596802401105171]
15. Nys G, Vanzandvoort M, Dekort P, Jansen B, Kappelle L, Dehaan E. Restrictions of the Mini-Mental State Examination in acute stroke. Archives of Clinical Neuropsychology 2005 Jul;20(5):623-629 [FREE Full text] [doi: 10.1016/j.acn.2005.04.001] [Medline: 15939186]
16. Mullick AA, Subramanian SK, Levin MF. Emerging evidence of the association between cognitive deficits and arm motor recovery after stroke: A meta-analysis. RNN 2015 Jun 17;33(3):389-403. [doi: 10.3233/rnn-150510]
17. Truong P, Lee J, Kwon A, Jeong G. Stride Counting in Human Walking and Walking Distance Estimation Using Insole Sensors. Sensors 2016 Jun 04;16(6):823. [doi: 10.3390/s16060823]
18. URL: www.R-project.org [accessed 2020-01-01]
19. David V, Forjan M, Martinek J, Kotzian S, Jagos H, Rafolt D. Evaluating wearable multimodal sensor insoles for motion-pattern measurements in stroke rehabilitation - A pilot study. IEEE Int Conf Rehabil Robot 2017 Jul;2017:1543-1548. [doi: 10.1109/ICORR.2017.8009467] [Medline: 28814039]
20. Kim SH, Kim JW, Park J, Shin MJ, Choi M. Predicting Energy Expenditure During Gradient Walking With a Foot Monitoring Device: Model-Based Approach. JMIR Mhealth Uhealth 2019 Oct 23;7(10):e12335. [doi: 10.2196/12335]
21. Bonnyaud C, Pradon D, Zory R, Bensmail D, Vuillerme N, Roche N. Gait parameters predicted by Timed Up and Go performance in stroke patients. NRE 2015 Feb 25;36(1):73-80. [doi: 10.3233/nre-141194]
22. 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]
23. Guzik A, Drużbicki M, Przysada G, Kwolek A, Brzozowska-Magoń A, Sobolewski M. Relationships between walking velocity and distance and the symmetry of temporospatial parameters in chronic post-stroke subjects. Acta Bioeng Biomech 2017;19(3):147-154 [FREE Full text] [Medline: 29205208]
24. Hiengkaew V, Jitaree K, Chaiyawat P. Minimal Detectable Changes of the Berg Balance Scale, Fugl-Meyer Assessment Scale, Timed “Up & Go” Test, Gait Speeds, and 2-Minute Walk Test in Individuals With Chronic Stroke With Different Degrees of Ankle Plantarflexor Tone. Archives of Physical Medicine and Rehabilitation 2012 Jul;93(7):1201-1208. [doi: 10.1016/j.apmr.2012.01.014]
25. Yu L, Xiong D, Guo L, Wang J. A remote quantitative Fugl-Meyer assessment framework for stroke patients based on wearable sensor networks. Computer Methods and Programs in Biomedicine 2016 May;128:100-110. [doi: 10.1016/j.cmpb.2016.02.012]

Abbreviations

FIM: functional independence measure
FMA: Fugl-Meyer Assessment
MBI: Modified Barthel Index
MMSE: Mini-Mental State Examination
TUG: Timed Up and Go

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by IV George; comments to author 29.07.20; revised version received 10.08.20; accepted 18.08.20; published 10.09.20.

Please cite as:
Seo M, Shin MJ, Park TS, Park JH
Clinometric Gait Analysis Using Smart Insoles in Patients With Hemiplegia After Stroke: Pilot Study
JMIR Mhealth Uhealth 2020;8(9):e22208
URL: http://mhealth.jmir.org/2020/9/e22208/
doi:10.2196/22208
PMID:32909949

Original Paper

Neural Network–Based Algorithm for Adjusting Activity Targets to Sustain Exercise Engagement Among People Using Activity Trackers: Retrospective Observation and Algorithm Development Study

Ramin Mohammadi[1,2], PhD; Mursal Atif[2], BA; Amanda Jayne Centi[2], PhD; Stephen Agboola[2,3,4], MD, PhD; Kamal Jethwani[2,3,4], MD, PhD; Joseph Kvedar[2,3,4], MD; Sagar Kamarthi[1,2], PhD

[1]Northeastern University, Boston, MA, United States

[2]Partners connected for health, Boston, MA, United States

[3]Massachusetts General Hospital, Department of Dermatology, Boston, MA, United States

[4]Harvard University, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Sagar Kamarthi, PhD
Northeastern University
360 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 6173733070
Email: sagar@coe.neu.edu

Abstract

Background: It is well established that lack of physical activity is detrimental to the overall health of an individual. Modern-day activity trackers enable individuals to monitor their daily activities to meet and maintain targets. This is expected to promote activity encouraging behavior, but the benefits of activity trackers attenuate over time due to waning adherence. One of the key approaches to improving adherence to goals is to motivate individuals to improve on their historic performance metrics.

Objective: The aim of this work was to build a machine learning model to predict an achievable weekly activity target by considering (1) patterns in the user's activity tracker data in the previous week and (2) behavior and environment characteristics. By setting realistic goals, ones that are neither too easy nor too difficult to achieve, activity tracker users can be encouraged to continue to meet these goals, and at the same time, to find utility in their activity tracker.

Methods: We built a neural network model that prescribes a weekly activity target for an individual that can be realistically achieved. The inputs to the model were user-specific personal, social, and environmental factors, daily step count from the previous 7 days, and an entropy measure that characterized the pattern of daily step count. Data for training and evaluating the machine learning model were collected over a duration of 9 weeks.

Results: Of 30 individuals who were enrolled, data from 20 participants were used. The model predicted target daily count with a mean absolute error of 1545 (95% CI 1383-1706) steps for an 8-week period.

Conclusions: Artificial intelligence applied to physical activity data combined with behavioral data can be used to set personalized goals in accordance with the individual's level of activity and thereby improve adherence to a fitness tracker; this could be used to increase engagement with activity trackers. A follow-up prospective study is ongoing to determine the performance of the engagement algorithm.

(JMIR Mhealth Uhealth 2020;8(9):e18142) doi:10.2196/18142

KEYWORDS

activity tracker; exercise engagement; dynamic activity target; neural network; activity target prediction; machine learning

Introduction

Background

Studies have reported the efficacy of physical activity in reducing the risk of disease; however, physical inactivity is on rise in the United States [1]. Considering that physical inactivity was the fourth leading cause of mortality in 2016 [2], there is much emphasis on developing effective methods to maintain healthy levels of physical activity. One promising solution is wearable fitness trackers that enable individuals to monitor their activity levels and patterns to ensure a healthy level of physical activity [3].

It is reported that about 20% of the general health-tracking population use smart devices such as medical gadgets, mobile phone apps, or online tools to track their health data [4]. The use of technology to objectively monitor physical activity is associated with higher levels of activity [5]. However, the potential benefits derived from the use of physical activity trackers are challenged by the limited and transient adoption of devices that necessarily require sustained use to exert their intended effect. Continued engagement with fitness trackers is an issue that warrants further investigation [5] A previous study [6] found two factors associated with the adoption and sustained use of physical activity trackers: (1) the number of digital devices owned by the participants, and (2) the use of activity fitness trackers and other smart devices by the participants' family members. The existence of these two factors bode well for the increased use of activity trackers; one study [1] demonstrated that motivational factors are associated with physical activity levels. Time constraints, fatigue, and aversion to exercise are some of the barriers to engaging in physical activity [1]. It has been reported that adjustments to activity targets are likely to enhance the users' commitment to physical activity and engagement with fitness trackers [7].

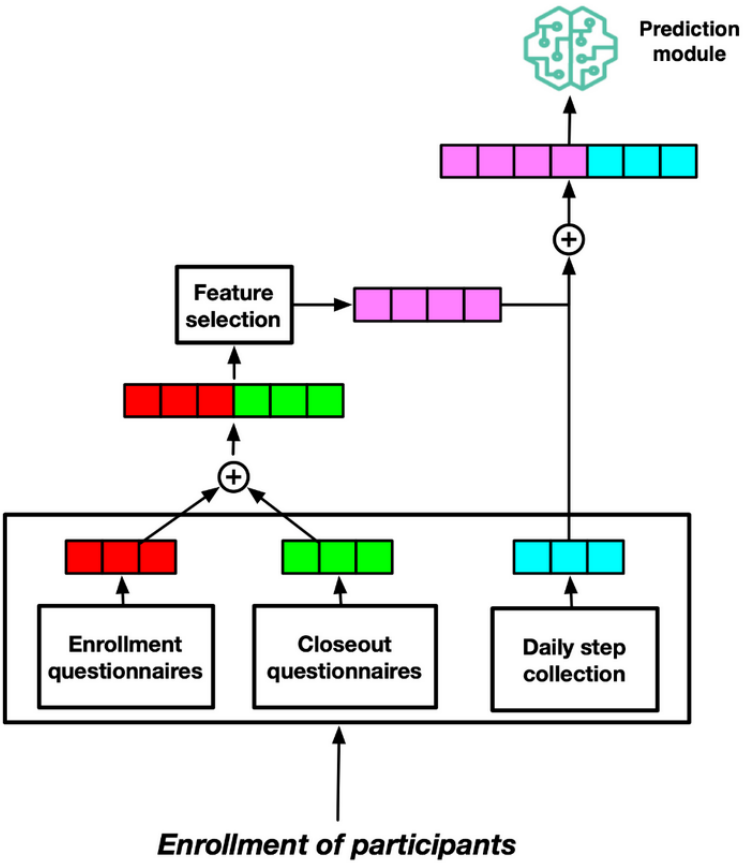
With the rise of machine learning and availability of activity tracker data, it is possible to create a model that can learn from users' behavior and adjust activity targets. Machine learning methods have broadly been applied to many health care areas such as cancer staging, risk assessment, and drug recommendation systems [8]. Researchers have studied the accuracy of activity trackers for energy expenditure assessment [9]. Having an automated personal trainer enabled by data mining techniques can be useful for an amateur athlete who cannot afford a personal trainer [10]. Similarly, effective feedback methods can be used for helping both athletes and coaches [11].

Although, there have been some attempts to study the benefit of activity trackers, there is room for studies on how to make use of activity trackers on a sustained basis. This study, to the best of our knowledge, is the first of its kind to develop a machine learning method to adjust the activity target for activity trackers. Machine learning techniques such as, but not limited to, lasso regression, ridge regression, Bayesian ridge regression, neural networks, random forest regression, and support vector regression have been used for prediction in the medical field [12].

Feature selection is an important step for improving model performance [13]. Prior to applying machine learning techniques, it is essential to study the data to find features that might negatively or positively affect the model [8]. In this work, we applied two feature selection techniques with a support vector machine [14]: principal component analysis [15] and recursive feature elimination.

We compared predictive models developed over (1) all features, (2) features generated by principal component analysis, (3) features selected by recursive feature elimination, and (4) features found from the authors' previous study [6] that characterizes participants environments. Figure 1 shows the study flow diagram highlighting the key steps undertaken in this study.

Figure 1. Study flow diagram.



Objective

The purpose of this study was to develop a predictive model to estimate achievable weekly step goals. By setting realistic targets that are neither too easy nor too difficult to achieve, activity tracker users can be encouraged to continue to meet their weekly activity goals, and at the same time, to find utility of the tracker. We chose individuals who were overweight as our first use case, because the benefits of sustaining or even increasing physical activity in this population are well known, while there have been a lack of sustained interventions addressing this issue [16-18].

Methods

Data Collection

The study was designed as a 9-week, nonrandomized pilot in which the data were analyzed retrospectively. For this purpose, adults (N=30) with a BMI of 25 kg/m^2 or greater were enrolled from a local Massachusetts General Hospital–affiliated clinic. After screening the participants and seeking their consent, the research team directed the study participants to visit the Wellocracy website [19] to read information regarding the study and review different types of activity trackers (and their features) available for their use during the 9-week study. The study staff assisted the participants with the device setup process as needed. For the study, 27 participants chose to use the FitBit Charge (Fitbit Inc) model, 2 chose the FitBit One (Fitbit Inc), and 1 chose the FitBit Zip (Fitbit Inc); 10 participants data were not used for analysis (7 participants either did not use the activity trackers at all or used less than 3 days a week, and an additional 3 participants demonstrated irregular use of their activity trackers on week by week basis). The data from the remaining 20 participants were used for model building.

The following surveys were collected from each participant both at enrollment and at the closeout stages: Behavioral Regulation in Exercise Questionnaire (BREQ-2) [20]; Barriers to Being Active (BBA) [21]; Patient Health Questionnaire (PHQ-8) [22]; Prochaska Stage of Change [23]; and Patient-Reported Outcomes Measurement Information System (PROMIS) Global-10 [24]. These surveys included questions about technology use and the ownership of electronic devices (BREQ-2), questions related to perceived barriers to exercise and activity (BREQ-2 and BBA), depression screening questions (PHQ-8), stages of change (Prochaska), and general health questions (PROMIS Global-10).

The BREQ-2 is designed to gauge the extent to which reasons for exercise are internalized and self-determined based on the following categories: motivation, external, introjected, identified, and intrinsic. In contrast, the BBA assesses whether participants gauge certain categories as reasons for inactivity and includes energy, willpower, time, and resources. A score of 5 or greater for a category indicates that it is a substantial barrier to a person's ability to exercise.

Participants were instructed to continuously wear the activity tracker for the entire period of the study. The first week was treated as a run-in period. Participants were contacted minimally during the remaining 8-week period to facilitate observation of participants' activity tracker habits without interference. At the

end of the study, participants completed a closeout survey either online or in paper format and underwent a phone interview to gather information on their experiences during the study. All interviews were conducted and transcribed for analysis by a trained neuropsychologist.

Experimental Setting

We divided the data into disjoint training and testing sets. We chose 16 participants' data as training data set and the remaining 4 participants' data were used as the testing data set. We explored and fine-tuned hyperparameters using the training data set. Each participant was informed and had a fixed weekly activity goal for weeks 2 through 8; it was fixed for each participant at 110% of their week 1 average daily step count. Since we didn't have a means to adjust the activity goal for each week for each participant, this value was used as an estimate of the personalized average daily step count goal for each week. We ignored the week if a participant used the tracker device for less than 3 days during the week.

Data Preprocessing

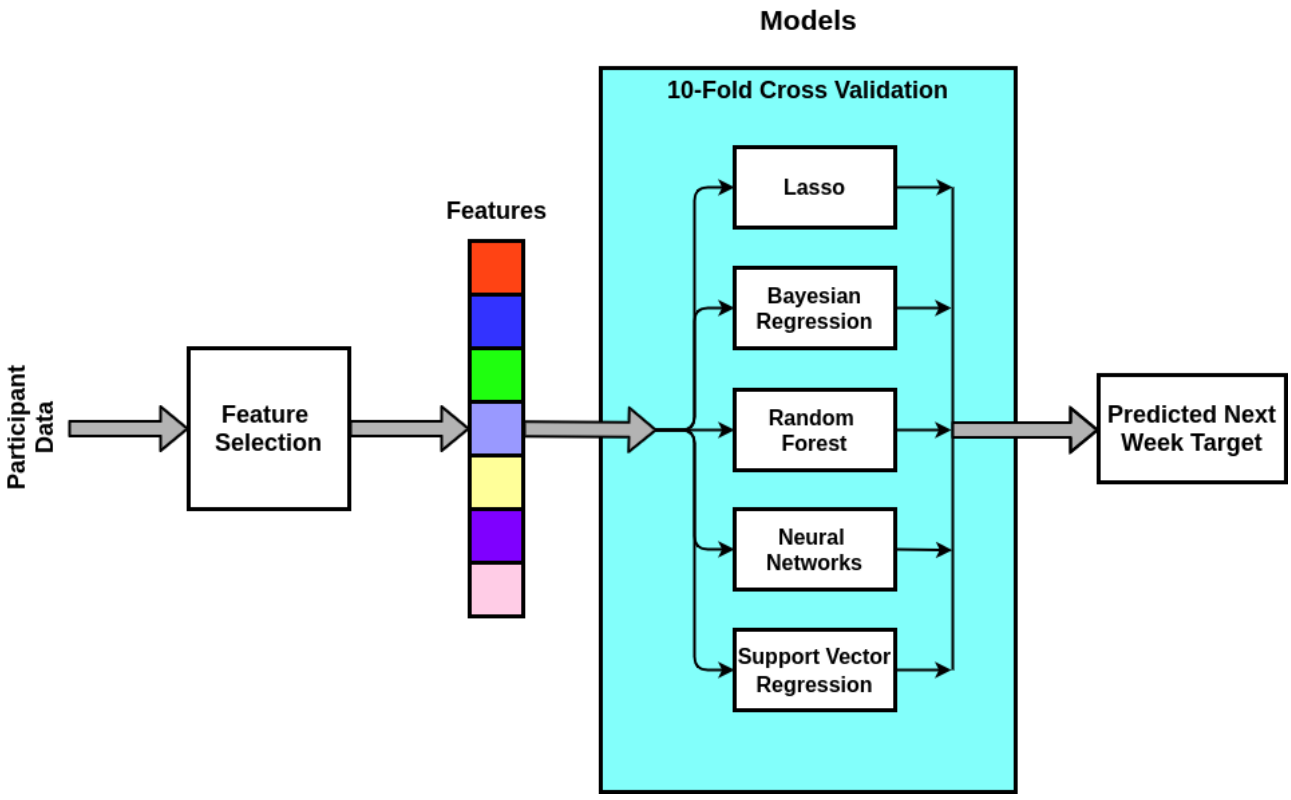
We collected all data from questionnaires and the participants' daily step count data from their activity trackers. The questionnaires generated 96 variables. The variables were screened to determine candidate predictors for building a machine learning model. In the first pass of the variable screening process, 11 variables whose variance was zero were eliminated. In the second pass, we examined pairwise correlations among all remaining variables to eliminate redundant ones. Since we found that all pairwise correlations were less than 0.6, we considered all variables to be nonredundant.

Machine Learning Techniques

Figure 2 presents the models used to predict participants' activity target. Selection of these models was based on their suitability and capabilities. Bayesian ridge regression estimates a probabilistic model of a ridge regression [25]. Lasso regression is robust to overfitting due to its regularization penalty [26]. Random forest models are versatile for numeric and categorical predictors and for classification and regression tasks. Random forest models can also be interpreted easily and are less susceptible to underfitting [27]. Neural network models consist of hundreds or thousands of neurons that perform mathematical operations to recognize patterns [28]. Support vector regression models are regression models whose optimization is unaffected by the dimensionality of the data [14].

We employed these models in four cases: (1) using all features, (2) using new features built by principal component analysis, (3) using important features found by recursive feature elimination, and (4) using a subset of features found from the authors' previous study [6] that characterizes participants environments.

Figure 2. A schematic depicting models developed using the training data.



Feature Selection and Extraction

In this study, we used participants' mental health, behavioral information, and weekly activity performance as inputs. For every 7-day moving window, the subsequent 7 days were considered as the forthcoming week. The questionnaires provided a highly redundant and low variance data set. These features coupled with 7 daily steps counts of the week and a normalized Shannon entropy (E_s) value of the weekly step count were considered candidate predictors for target step count for

each individual participant for the forthcoming week [29]. The normalized Shannon entropy was computed as  where i denotes the day, p_i denotes the portion of the total weekly steps completed on day i, and N denotes the number of days per week (ie, 7). The normalized Shannon entropy varies between 0 and 1. A value close to 0 indicates that the daily step counts of the participant throughout the week were irregular; in contrast, a value close 1 indicates that the step counts were consistent.

In total there were 93 candidate features for building the predictive models: 85 features from the questionnaires, 7 features from the daily steps of the week, and one weekly feature (ie, Shannon entropy). We used all 93 features for models developed in case 1.

For case 2, we developed a principal component analysis model. Principal component analysis is a dimension reduction technique that is widely used for extracting uncorrelated features components from correlated variables [14]. The principal component analysis used the 85 features from the questionnaires. We combined the principal components with the 7 daily and 1 weekly step count features.

In case 3, to identify the important features, we performed recursive feature elimination with support vector regression using the 85 questionnaire variables. We augmented these features with the 7 daily step count features and the normalized Shannon entropy value of the weekly step count.

Lastly, for case 4, we selected 2 features that were found to be important from the questionnaire in our previous work [6] studying the link between the participants' environmental factors and the adherence to the use of activity trackers. These features were (1) the number of digital devices owned by the participants, and (2) the use of activity fitness trackers and other smart devices by the participants' family members. We appended these two features with the daily step count features, and Shannon entropy of the weekly step count.

Model Evaluation

We trained the models in this study with 10-fold cross-validation using the training data set. We compared the performance of the models using mean absolute error (MAE) and adjusted R^2. We tested the final models (models A, B, C and D) of each case (cases 1, 2, 3, and 4) on an unseen test data set.

Statistical Analysis

We used Python (version 3.5.0) and R (version 3.4.1) for model development and statistical tests. We performed a one-sided paired t test ($P<.05$ were deemed significant) for statistical comparison between the models. The null hypothesis was that the mean error of two models were equal, with the alternative hypothesis that the candidate model had a lower mean error than the mean error given by the comparison model.

Results

Participant Characteristics

We present the characteristics of the study participants in Table 1. Among them, 10 participants were eventually removed from the study since they stopped using the activity trackers.

Over the span of 8 weeks, not all participants met their week-by-week step objectives. By and large, under half of participants met their progression objective every week (see Table 2). We present the distribution step count for all participants for each of the 9 weeks in Figure 3. We also presented the distribution of steps for each day of the week during the 9-week study period (see Figure 4).

Table 1. Patient demographic data.

Variable	Enrolled (N=30)	Participants (n=20)
Age (years), mean (SD)	48.9 (9.5)	47.7 (10.2)
Gender, n (%)		
Male	9 (30)	6 (30)
Female	21 (70)	14 (70)
BMI at enrollment		
Mean (SD)	32.5 (4.6)	32.8 (4.7)
Range	25.0-41.2	25.0-41.2
Race, n (%)		
White	21 (70)	14 (70)
American Indian or Alaskan Native	1 (3)	1 (5)
Black or African American	3 (10)	2 (10)
Hispanic	3 (10)	3 (15)
Unknown	2 (6)	0 (0)
Marital status, n (%)		
Married	8 (27)	6 (30)
Divorced or separated	8 (27)	5 (25)
Single (never married)	8 (27)	6 (30)
Living with partner	3 (10)	2 (10)
Widowed	1 (3)	1 (5)
No response	2 (7)	0 (0)
Education, n (%)		
12 years or completed high school or GED	5 (17)	3 (15)
Some college	5 (17)	2 (10)
College graduate	9 (30)	8 (40)
Post–high school	2 (7)	2 (10)
Postgraduate	2 (7)	1 (5)
Less than high school	3 (10)	2 (10)
Unknown	4 (13)	1 (5)
Employment status, n (%)		
Employed/self-employed	15 (50)	12 (60)
Disabled	5 (17)	3 (15)
Unemployed	5 (17)	2 (10)
Student	1 (3)	1 (5)
Retired	1 (3)	1 (5)
Unknown	3 (10)	1 (5)

Table 2. Participants meeting their average daily step goal for the week (110% of the average daily step count in week 1) over the course of the study.

Week	Participants (n=20) who met goal, n (%)
2	4 (20)
3	10 (50)
4	9 (45)
5	4 (20)
6	8 (40)
7	6 (30)
8	4 (20)
9	5 (25)

Figure 3. Weekly distribution of steps among all 20 participants.

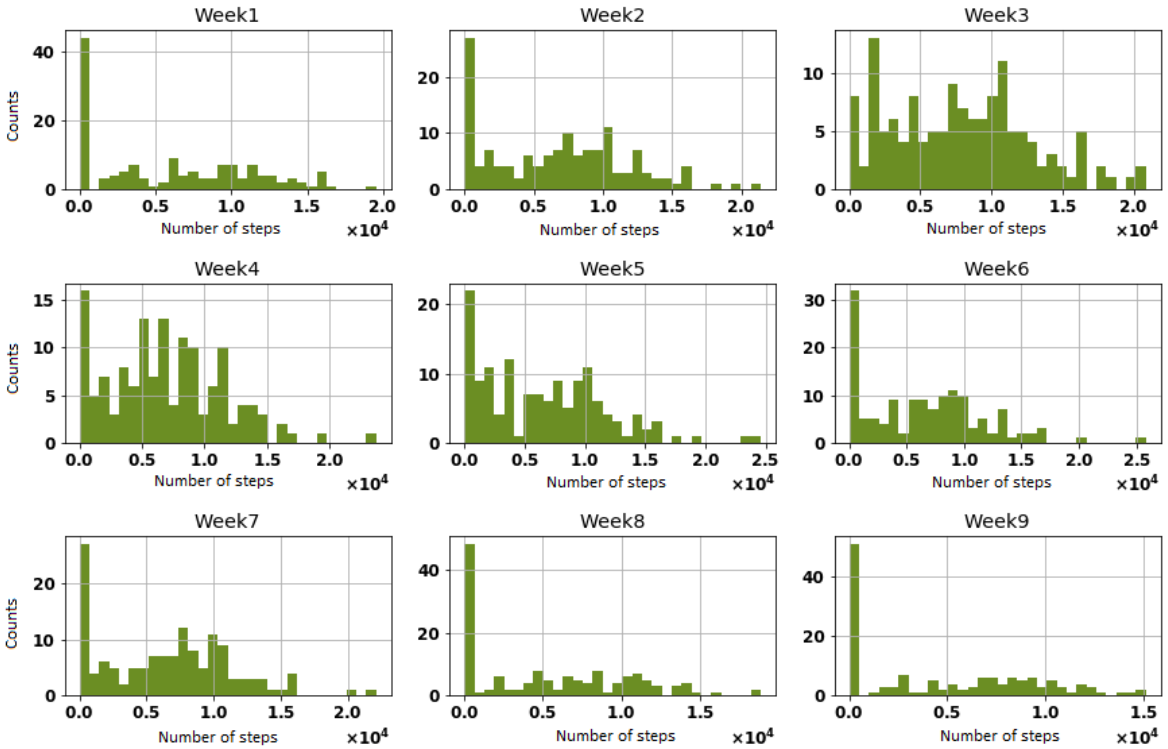
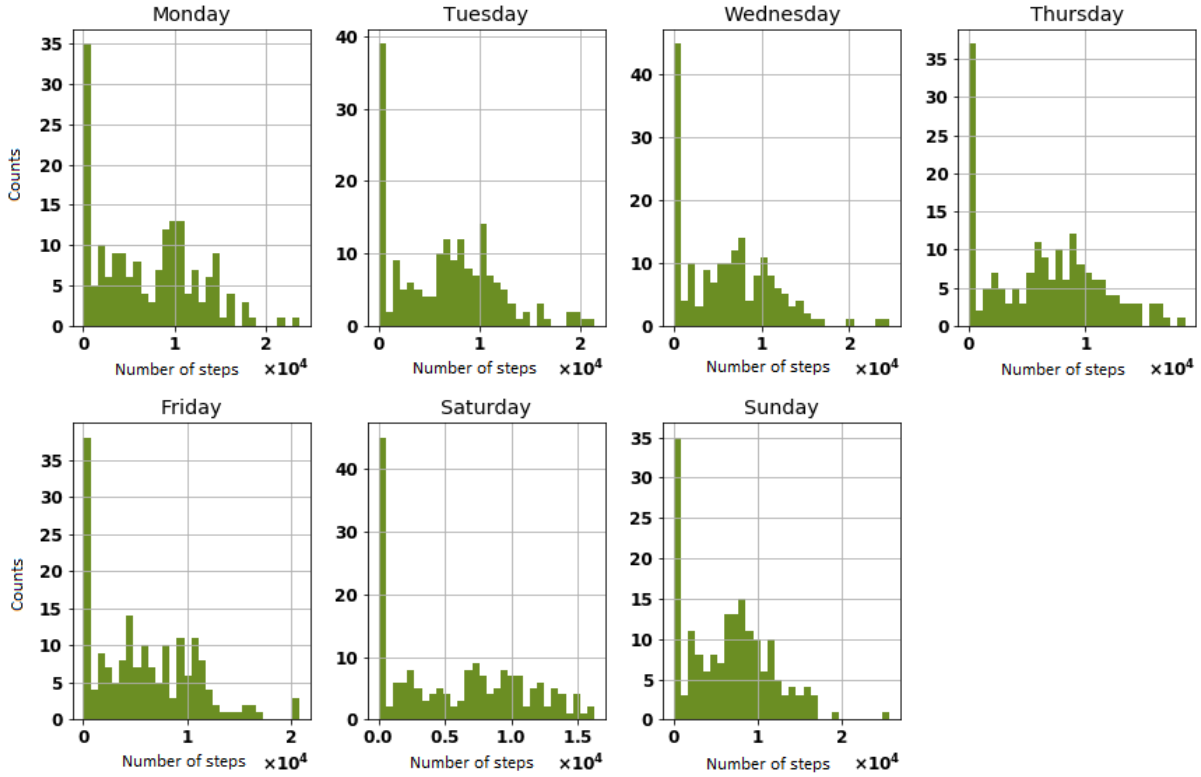


Figure 4. Step count distribution on each day of the week.



Parameter Tuning and Feature Selection

We performed a grid-search with 10-fold cross validation over the training data set. Lasso regression with α=0.01, Bayesian ridge regression with α=0.0001, and support vector regression with polynomial kernel of degree 2 with a regularization parameter of 0.00001 gave the best performance. Optimal parameters for random forests and neural networks depend on feature selection techniques. We found 30 variables from the questionnaires to be important using recursive feature elimination with support vector regression (see Figure 5). In the case of principal component analysis, the top 14 principal components explained 100% of the variance of the variables in questionnaires as shown in Figure 6.

Figure 5. Number of features found to be important by support vector regression.

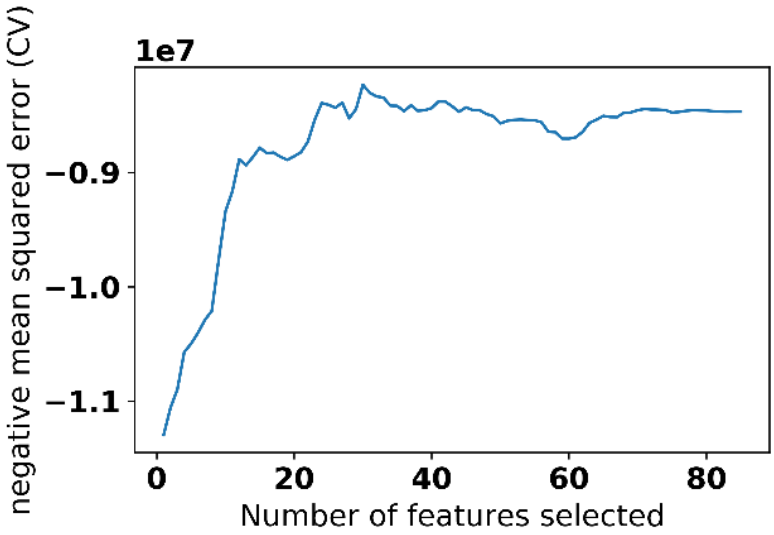
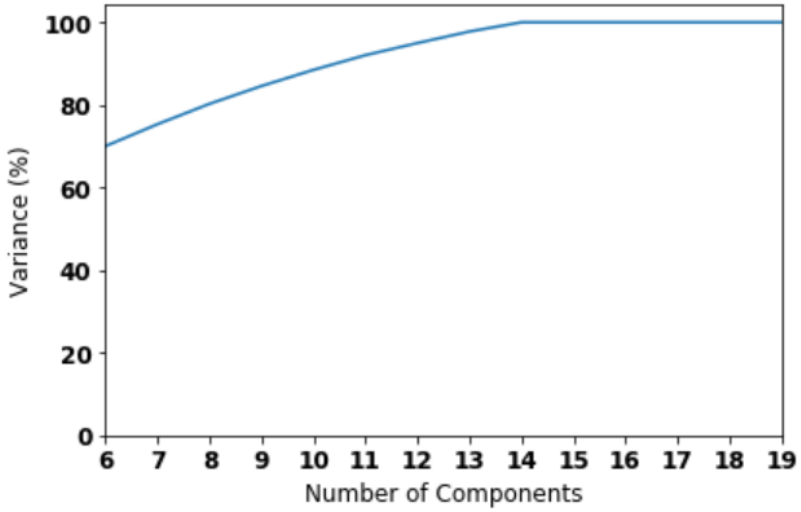


Figure 6. Number of components found by principal component analysis.



Model Performance

We report the performance for all models of 10-fold cross-validation with the training data set in Table 3. We used mean absolute error (MAE) and adjusted R^2 for the model comparison. Among all models developed, the Bayesian ridge regression model that used 93 features (case 1) gave the best performance for the training set (see Figure 7; MAE 1672, 95% CI 1640-1704; adjusted R^2=0.85). The Bayesian ridge regression model is referred to as model A in the rest of this paper.

Similarly, in case 2, we appended 14 principal components extracted from principal component analysis to the 7 daily and 1 weekly step count features resulting in 22 features. A random forest model (MAE 1700, 95% CI 1664-1737; adjusted R^2=0.91) gave the best performance among all models (see Figure 8). This random forest model is referred to as model B for the rest of this paper.

Table 3. Results of 10-fold cross-validation with training set.

Case and models	Mean MAE	95% CI	*P* value
1 All features			
Bayesian ridge regression[a]	1672	1640-1704	Model ref[b]
Lasso	2016	1985-2047	.002
Random forest	2425	2386-2464	.002
Neural network	1856	1813-1899	.002
Support vector regression	2425	2386-2464	.002
2 Feature selection using principal component analysis			
Bayesian ridge regression	2139	2107-2171	<.001
Lasso	2036	2005-2067	<.001
Random forest[c]	1700	1664-1737	Model ref
Neural network	1956	1926-1985	.03
Support vector regression	2938	2862-3013	<.001
3 Feature selection using recursive feature elimination			
Bayesian ridge regression	2131	2100-2163	.002
Lasso	2026	1995-2057	.002
Random forest[d]	1774	1739-1809	Model ref
Neural network	1906	1855-1958	.002
Support vector regression	2548	2473-2624	.002
4 Feature selected from previous study			
Bayesian ridge regression	2564	2537-2592	<.001
Lasso	2457	2429-2485	<.001
Random forest	1810	1768-1852	.04
Neural network[e]	1622	1589-1655	Model ref
Support vector regression	2940	2864-3016	<.001

[a]Model A (ie, this model gave the best performance in case 1).

[b]Reference model for comparisons.

[c]Model B (ie, this model gave the best performance in case 2).

[d]Model C (ie, this model gave the best performance in case 3).

[e]Model D (ie, this model gave the best performance in case 4).

Figure 7. Cross-validation performance of models using all questionnaire features, 7 daily step count features, and weekly entropy feature. BRIDGE: Bayesian ridge regression; NN: neural network; RF: random forest; SVR: support vector regression.

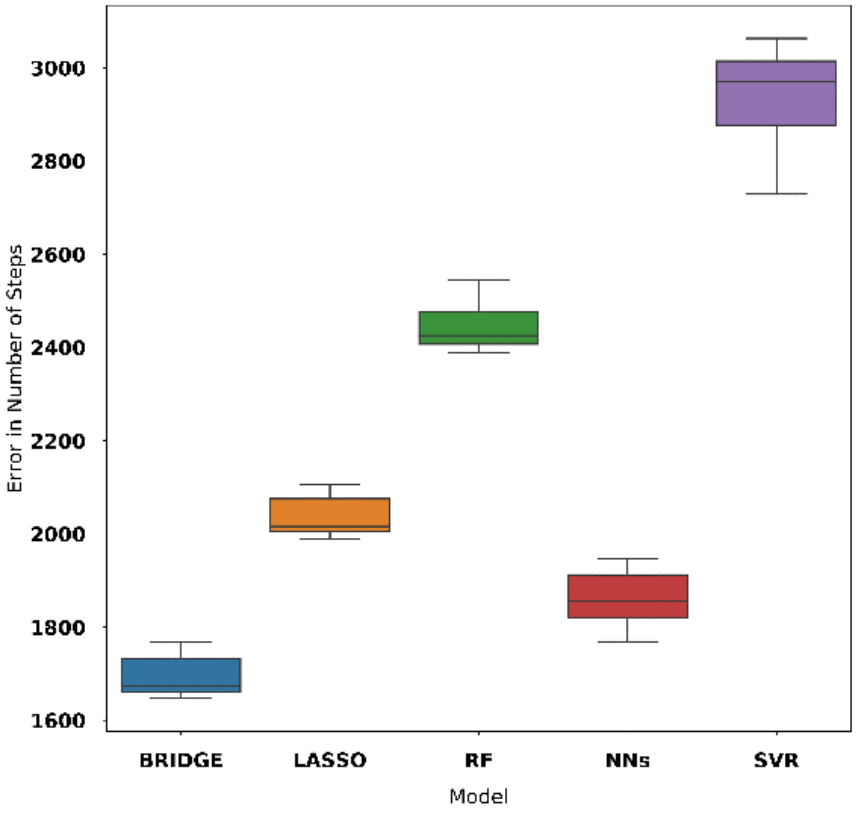
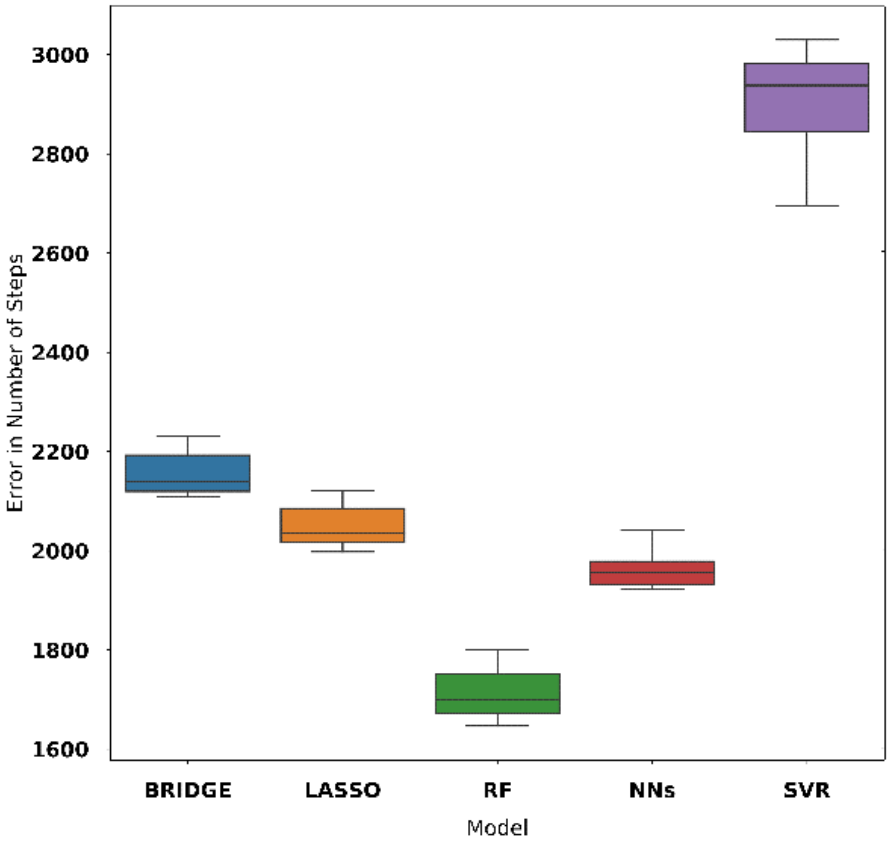


Figure 8. Cross-validation performance of models using features generated by principal component analysis, daily step count features, and weekly entropy feature. BRIDGE: Bayesian ridge regression; NN: neural network; RF: random forest; SVR: support vector regression.



In case 3, we appended variables found by recursive feature elimination to the 7 daily step counts and 1 weekly entropy feature which resulted in 39 features. A random forest model (MAE 1774, 95% CI 1739-1809; adjusted R^2=0.81) offered the best performance among all the models developed with these features (see Figure 9). We refer to this random forest model as model C.

Finally, for case 4, we coupled two features found in the previous study [6] with the 7 daily step counts and 1 weekly entropy feature which resulted in 10 features. A neural network gave the best performance across all models (MAE 1622 steps, 95% CI 1589-1655; adjusted R^2=0.89) (see Figure 10). We refer to this neural network model as model D.

Model D gave the best predictive performance among all the models. It had the lowest MAE across all models explored. We compared the predictive power of the model D (neural networks) with those of models A, B, and C using the testing data set. We found that model D gave a better predictive performance as shown in Figure 11. We performed comparisons using *t* tests between the errors generated by the model D and errors generated by models A, B, and C as shown in Table 4. We observed that model D's lower errors in comparison to those of Bayesian ridge regression (model A: *P*=.01), random forest (model B: *P*<.001), and random forest (model C: *P*=.01) models were significant.

Figure 9. Cross-validation performance of models using all features given by recursive feature elimination, 7 daily step count features, and weekly entropy feature. BRIDGE: Bayesian ridge regression; NN: neural network; RF: random forest; SVR: support vector regression.

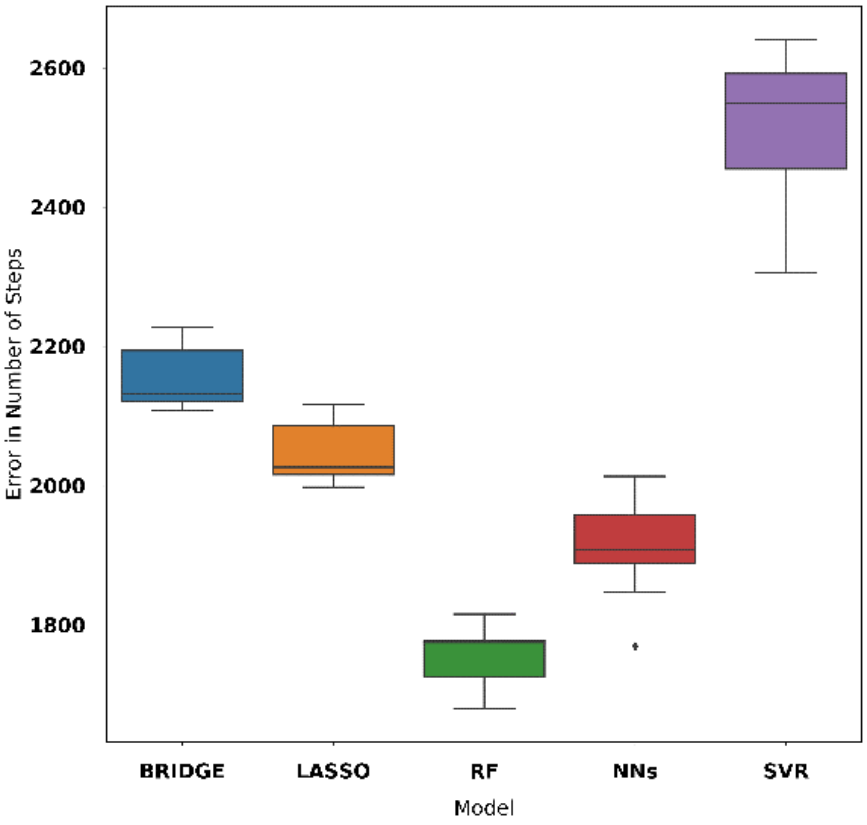


Figure 10. Cross-validation performance of models using features generated from previous knowledge, 7 daily step count features, and weekly entropy feature. BRIDGE: Bayesian ridge regression; NN: neural network; RF: random forest; SVR: support vector regression.

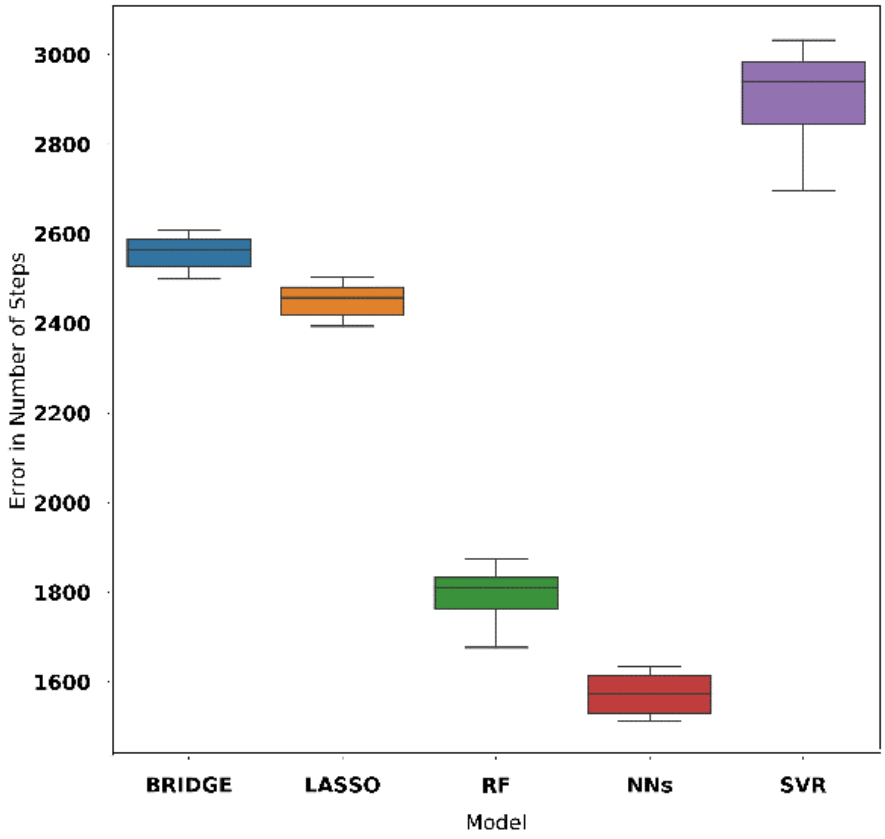


Figure 11. Boxplot of errors in terms of steps over the test set for Model A (Bayesian ridge regression), Model B (random forest), Model C (random forest), and Model D (neural network). MAE: mean absolute error.

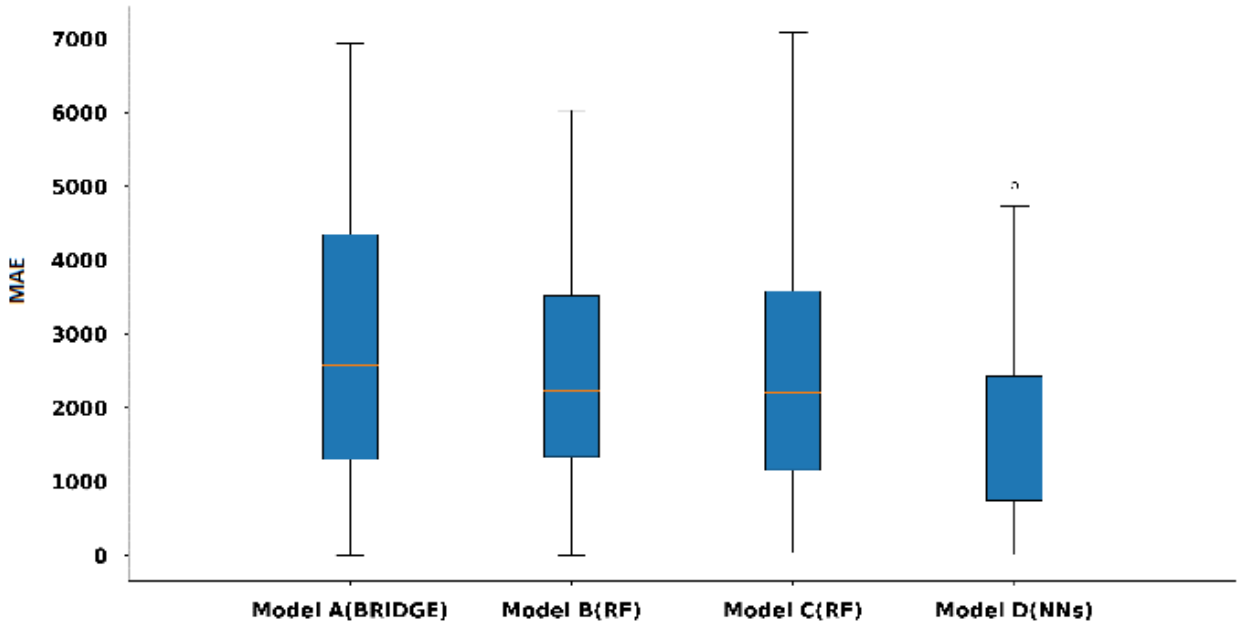


Table 4. Breakdown of model results over the test data set.

Model (type)	Mean MAE[a]	95% CI	*P* value
Model D (neural network)	1545	1383-1706	Model ref[b]
Model C (random forest)	2210	1990-2420	.01
Model B (random forest)	2230	2015-2445	<.001
Model A (Bayesian ridge regression)	2578	2310-2845	.01

[a]MAE: mean absolute error.

[b]Reference model for comparisons.

We evaluated the naïve rule: using participants' average daily step count of the week as the prediction for the subsequent week's activity goal [30]. This is a reasonably competitive approach because the weekly target exhibited strong autocorrelation. The naïve rule achieved an MAE of 1664 steps while model D achieved an MAE of 1545 steps (95% CI 1383-1706) for the 4 test participants over a period of 8 weeks.

Sensitivity Analysis

Its reported that the Fitbit activity devices have margin of ±5% error in their step count readings [31,32]. We tested the performance of our best model (model D) on three noisy data sets generated from the original test data set by adding ±1%, ±3%, and ±5% noise. The first noisy data set was generated by adding random noise between –1% and +1% to the Fitbit readings. Similarly, the second and third noisy data sets were generated by adding ±3% and ±5% noise to the original data set. To conduct the sensitivity analysis, we generated 100 replications of each noisy data sets by adding different random noises with limit specified for each data set. These data sets were used to evaluate the performance of model D.

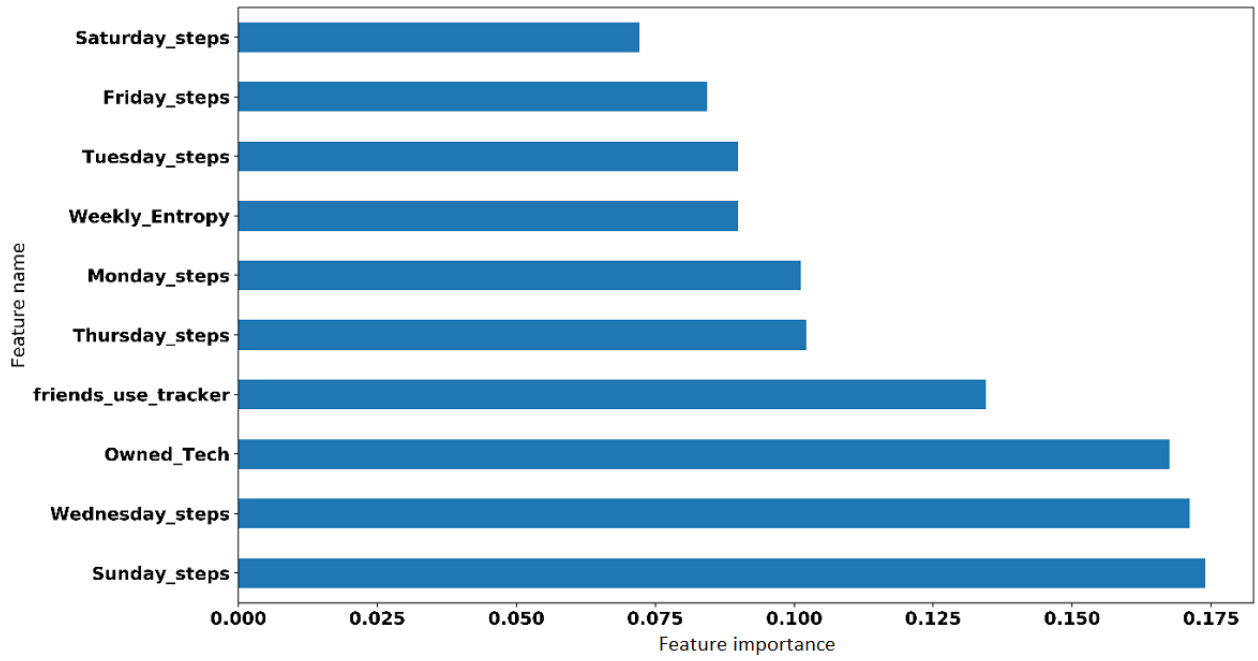
The model D achieved an MAE of 1606 steps (95% CI 1490-1755) for the noisy test data set generated using margin of ±1% error which is approximately 4% higher than the MAE achieved for the original noise-free test data set. Similarly, model D's MAE was 1670 steps (95% CI 1571-1840) for the margin of ±3% error which was approximately 8% higher than the MAE achieved for the original noise-free test data set. Finally, for the data set generated using margin of ±5% error, model D's MAE of 1710 steps (95% CI 1621-1908) was approximately 10% higher than the MAE achieved for the original noise-free test data set.

This led to the empirical observation that 2%, 6%, and 10% noise in Fitbit readings leads to approximately 4%, 8%, and 10% increases in prediction error, respectively. Therefore, we assume that the model error sublinearly increases with margin of error in activity tracker readings. However, one might need to validate this observation with further evaluation data.

Feature Importance

We experimented with integrated gradients [33] in order to analyze the features of model D. This method provides a score that reflects the contribution of each variable to the response variable by calculating the integral of the gradient of the response variable with respect to that variable. We report the top features for model D in Figure 12.

Figure 12. Importance of the features as measured by the integrated gradient method.



Discussion

Precise prescient calculations that consolidate information are one of the main focuses in predictive analyses [34]. To the best of our knowledge, this work is one of the first studies to explore machine learning models with the aim of adjusting step count goals. Inputs to these models use an individual's personal, social, and environmental factors as well as weekly activity data. A recent study concluded that activity tracker users feel unmotivated despite having knowledge about the benefits of physical activity [35]. Users can become unmotivated if they cannot meet their activity goal [36]. Step count goals that are too easy to achieve may lead to abandonment of the activity tracker, and those that are too high will discourage the individual [37].

In previous work [6], we studied the factors influencing the use of activity trackers and identified two factors that likely promote the continued use of activity trackers: (1) the number of digital devices that the participant owns, and (2) whether or not members of the participants' family use activity fitness trackers and other smart devices. Extending the previous work [6], in this study, we explored different predictive models to estimate achievable weekly step goals to encourage the use of activity trackers. This study can be used to set goals for individuals and can be accompanied by proper motivation messages to improve the sustained use of the activity trackers [38-40].

This study has some limitations. The number of participants was low, and all of the participants were selected from a cohort with BMI of 25 kg/m^2 or greater from the same geographical area. Participants who choose to participate in this study were more likely to use from 3 models of Fitbit activity trackers than those who chose not to participate in the study. Moreover, only participants who completed the 9-week study were considered for further analysis. Finally, in this study, we used 110% of week 1 averaged daily step count as the best estimate of each participant's personalized goal for each week. The goal may have been on heavy side for some participants and on the easy side for others. This is, of course, a limitation of the study in the absence of a mechanism for a correct estimate of the goal for each participant.

As an extension to this work, a new study with the goal of validating the developed model with a large number of participants has been undertaken. The new study recruited 120 individuals from the general public to use the neural network-based predictive model developed herein over a period of 6 months. This model, hosted on a server, provides each user with achievable daily steps goals and updates the model parameters on a weekly basis. The data from the study that is underway will be used to fine tune the predictive model and to gain insight into activity tracker users to motivate and manage their physical activity.

Acknowledgments

This research was supported in conjunction with a grant from the Robert Wood Johnson Foundation (Grant ID: 71963).

Authors' Contributions

RM and SK worked on data exploration, dimension reduction through factor analysis, implementing and designing the neural network–based algorithm, generating results, and drafting the paper. SK is one of the guarantor authors of the paper. MA and AJC worked on the design and data collection. SA and KJ worked on the design, and JK is one of the guarantor authors.

Conflicts of Interest

None declared.

References

1. Bassuk SS, Manson JE. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. Journal of Applied Physiology 2005 Sep;99(3):1193-1204. [doi: 10.1152/japplphysiol.00160.2005]
2. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior Change Techniques Present in Wearable Activity Trackers: A Critical Analysis. JMIR mHealth uHealth 2016 Apr 27;4(2):e40. [doi: 10.2196/mhealth.4461]
3. Eheman C, Henley SJ, Ballard-Barbash R, Jacobs EJ, Schymura MJ, Noone A, et al. Annual Report to the Nation on the status of cancer, 1975-2008, featuring cancers associated with excess weight and lack of sufficient physical activity. Cancer 2012 Mar 28;118(9):2338-2366. [doi: 10.1002/cncr.27514]
4. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: A meta-regression. Health Psychology 2009;28(6):690-701. [doi: 10.1037/a0016136]
5. Wang JB, Cataldo JK, Ayala GX, Natarajan L, Cadmus-Bertram LA, White MM, et al. Mobile and Wearable Device Features that Matter in Promoting Physical Activity. JournalMTM 2016 Jul;5(2):2-11. [doi: 10.7309/jmtm.5.2.2]
6. Centi AJ, Atif M, Golas SB, Mohammadi R, Kamarthi S, Agboola S, et al. Factors Influencing Exercise Engagement When Using Activity Trackers: Nonrandomized Pilot Study. JMIR Mhealth Uhealth 2019 Oct 24;7(10):e11603 [FREE Full text] [doi: 10.2196/11603]
7. Rúben G, Evangelos K, Marc H. How do we engage with activity trackers?: a longitudinal study of habito. 2015 Sep 07 Presented at: ACM International Joint Conference on Pervasive Ubiquitous Computing; 2015; New York p. 1305-1316. [doi: 10.1145/2750858.2804290]

8. Koh HC, Tan G. Data mining applications in healthcare. J Healthc Inf Manag 2005;19(2):64-72. [Medline: 15869215]
9. Phelan S, Phipps MG, Abrams B, Darroch F, Grantham K, Schaffner A, et al. Does behavioral intervention in pregnancy reduce postpartum weight retention? Twelve-month outcomes of the Fit for Delivery randomized trial. Am J Clin Nutr 2014 Feb;99(2):302-311 [FREE Full text] [doi: 10.3945/ajcn.113.070151] [Medline: 24284438]
10. Fister, Iztok, Duan F, Simon F. Data Mining in Sporting Activities Created by Sports Trackers. 2013 Aug 26 Presented at: International Symposium on Computational and Business Intelligence. IEEE; 2013; India. [doi: 10.1109/iscbi.2013.25]
11. Liebermann DG, Katz L, Hughes MD, Bartlett RM, McClements J, Franks IM. Advances in the application of information technology to sport performance. Journal of Sports Sciences 2002 Jan;20(10):755-769. [doi: 10.1080/026404102320675611]
12. Tomar D, Agarwal S. A survey on Data Mining approaches for Healthcare. IJBSBT 2013 Oct 31;5(5):241-266. [doi: 10.14257/ijbsbt.2013.5.5.25]
13. Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering 2014 Jan;40(1):16-28. [doi: 10.1016/j.compeleceng.2013.11.024]
14. Drucker, Harris. Support vector regression machines. Advances in neural information processing systems. Jolliffe, Principal component analysis, Springer 1997:155-161.
15. Jolliffe IT. Principal Component Analysis, 2nd ed. New York: Springer; 2002.
16. Söderlund A, Fischer A, Johansson T. Physical activity, diet and behaviour modification in the treatment of overweight and obese adults: a systematic review. Perspect Public Health 2009 May 07;129(3):132-142. [doi: 10.1177/1757913908094805]
17. Jakicic JM, Otto AD, Lang W, Semler L, Winters C, Polzien K, et al. The Effect of Physical Activity on 18-Month Weight Change in Overweight Adults. Obesity 2010 Jun 10;19(1):100-109. [doi: 10.1038/oby.2010.122]
18. Bickmore T, Schulman D, Yin L. Maintaining Engagement in Long-term Interventions with Relational Agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
19. http://www.wellocracy.com. URL: http://www.wellocracy.com [accessed 2020-07-28]
20. David M, Vannessa T. A modification to the Behavioural Regulation in Exercise Questionnaire to include an assessment of amotivation. Journal 2004:191-196. [doi: 10.1123/jsep.26.2.191]
21. Overcoming Barriers to Physical Activity. website. URL: https://www.cdc.gov/physicalactivity/basics/adding-pa/barriers.html [accessed 2020-07-28]
22. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. Journal of Affective Disorders 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026]
23. Holmen H, Wahl A, Torbjørnsen A, Jenum AK, Småstuen MC, Ribu L. Stages of change for physical activity and dietary habits in persons with type 2 diabetes included in a mobile health intervention: the Norwegian study in RENEWING HEALTH. BMJ Open Diab Res Care 2016 May 12;4(1):e000193. [doi: 10.1136/bmjdrc-2016-000193]
24. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009 Jun 19;18(7):873-880. [doi: 10.1007/s11136-009-9496-9]
25. Christopher M B. Pattern recognition and machine learning. In: springer. New York, NY: Springer; 2006.
26. Santosa F, Symes WW. Linear Inversion of Band-Limited Reflection Seismograms. SIAM J. Sci. and Stat. Comput 1986 Oct;7(4):1307-1330. [doi: 10.1137/0907087]
27. Genuer R, Poggi J, Tuleau-Malot C. Variable selection using random forests. Pattern Recognition Letters 2010 Oct;31(14):2225-2236. [doi: 10.1016/j.patrec.2010.03.014]
28. Simon H. Neural networks: a comprehensive foundation. In: book. NJ, United States: Prentice Hall PTR; 1994.
29. Cover, Thomas M. ,Joy A. Thomas.Elements of information theory. John Wiley & Sons 2012.
30. Mohammadi, Jain S, Agboola S, Palacholla R, Kamarthi S, Wallace BC. Learning to Identify Patients at Risk of Uncontrolled Hypertension Using Electronic Health Records Data. AMIA Jt Summits Transl Sci Proc 2019;2019:533-542 [FREE Full text] [Medline: 31259008]
31. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities. JMIR Mhealth Uhealth 2017 Aug 11;5(8):e106. [doi: 10.2196/mhealth.6321] [Medline: 28801304]
32. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
33. Sundararajan, Mukund, Ankur T, Qiqi Y. Axiomatic attribution for deep networks. 2017 Aug 11 Presented at: Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. Org; 2017; Sydney, Australia.
34. Rykov Y, Thach T, Dunleavy G, Roberts AC, Christopoulos G, Soh CK, et al. Activity Tracker–Based Metrics as Digital Markers of Cardiometabolic Health: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 Jan 27;8(1):e16409. [doi: 10.2196/16409]
35. Glenn F, Shaun G, Andrew K, Leslie S. Why People Stick With or Abandon Wearable Devices. Catalyst Journal 2017 Sep 14.

36. Kerner C, Goodyear VA. The Motivational Impact of Wearable Healthy Lifestyle Technologies: A Self-determination Perspective on Fitbits With Adolescents. American Journal of Health Education 2017 Jul 28;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
37. Elizabeth M, David H, Gueorgi K. Mobile health apps: adoption, adherence, and abandonment. In: UbiComp/ISWC'15 Adjunct: Adjunct Proceedings. 2015 Sep 14 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings; 2015; Osaka Japan p. 261-264. [doi: 10.1145/2800835.2800943]
38. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and Retention: Measuring Breadth and Depth of Participant Use of an Online Intervention. J Med Internet Res 2010 Nov 18;12(4):e52. [doi: 10.2196/jmir.1430]
39. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a Web-Based Physical Activity Intervention for Patients With Knee and/or Hip Osteoarthritis: A Mixed Method Study. J Med Internet Res 2013 Oct 16;15(10):e223. [doi: 10.2196/jmir.2742]
40. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective Associations Between Intervention Components and Website Engagement in a Publicly Available Physical Activity Website: The Case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4. [doi: 10.2196/jmir.1792]

Abbreviations

BBA: Barriers to Being Active
BREQ: Behavioral Regulation in Exercise Questionnaire
MAE: mean absolute error
PHQ: Patient Health Questionnaire
PROMIS: Patient-Reported Outcomes Measurement Information System

Edited by G Eysenbach; submitted 06.02.20; peer-reviewed by M Pijl, K Hino; comments to author 23.03.20; revised version received 26.05.20; accepted 03.06.20; published 08.09.20.

Please cite as:
Mohammadi R, Atif M, Centi AJ, Agboola S, Jethwani K, Kvedar J, Kamarthi S
Neural Network–Based Algorithm for Adjusting Activity Targets to Sustain Exercise Engagement Among People Using Activity Trackers: Retrospective Observation and Algorithm Development Study
JMIR Mhealth Uhealth 2020;8(9):e18142
URL: https://mhealth.jmir.org/2020/9/e18142
doi:10.2196/18142
PMID:32897235

Corrigenda and Addenda

Correction: mHealth Interventions to Promote Anti-Retroviral Adherence in HIV: Narrative Review

Stephen B Lee[1,2], MD, FRCPC; Joanne Valerius[2], PhD

[1]Department of Medicine, Division of Infectious Diseases, University of Saskatchewan College of Medicine, Regina, SK, Canada
[2]Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR, United States

Corresponding Author:
Stephen B Lee, MD, FRCPC
Department of Medicine, Division of Infectious Diseases
University of Saskatchewan College of Medicine
Regina General Hospital, Unit 4E
Regina, SK, S4P 0W5
Canada
Phone: 1 306 766 4247
Email: leestephenz@gmail.com

Related Article:

Correction of: https://mhealth.jmir.org/2020/8/e14739/

(JMIR Mhealth Uhealth 2020;8(9):e24250) doi:10.2196/24250

In "mHealth Interventions to Promote Anti-Retroviral Adherence in HIV: Narrative Review" (JMIR Mhealth Uhealth 2020;8(8):e14739) the authors noted one error.

In the originally published manuscript, 2nd author Joanne Valerius was noted as having contributed equally. This was incorrect and this note has now been removed.

The correction will appear in the online version of the paper on the JMIR Publications website on September 15, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 10.09.20; this is a non–peer-reviewed article;accepted 10.09.20; published 15.09.20.

Please cite as:
Lee SB, Valerius J
Correction: mHealth Interventions to Promote Anti-Retroviral Adherence in HIV: Narrative Review
JMIR Mhealth Uhealth 2020;8(9):e24250
URL: http://mhealth.jmir.org/2020/9/e24250/
doi:10.2196/24250
PMID:32931442

Original Paper

Digital Cardiovascular Biomarker Responses to Transcutaneous Cervical Vagus Nerve Stimulation: State-Space Modeling, Prediction, and Simulation

Asim H Gazi[1], BS; Nil Z Gurel[1], MS; Kristine L S Richardson[1], BS; Matthew T Wittbrodt[2], PhD; Amit J Shah[3,4,5], MPH, MD; Viola Vaccarino[3,4], MD, PhD; J Douglas Bremner[2,5,6], MD; Omer T Inan[1,7], PhD

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, United States

[2]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, United States

[3]Department of Epidemiology, Rollins School of Public Health, Atlanta, GA, United States

[4]Department of Medicine, Division of Cardiology, Emory University School of Medicine, Atlanta, GA, United States

[5]Atlanta VA Medical Center, Emory University, Atlanta, GA, United States

[6]Department of Radiology, Emory University School of Medicine, Atlanta, GA, United States

[7]Coulter Department of Bioengineering, Georgia Institute of Technology, Atlanta, GA, United States

Corresponding Author:
Asim H Gazi, BS
School of Electrical and Computer Engineering
Georgia Institute of Technology
North Ave NW
Atlanta, GA, 30332
United States
Phone: 1 4693608083
Email: asim.gazi@gatech.edu

Abstract

Background: Transcutaneous cervical vagus nerve stimulation (tcVNS) is a promising alternative to implantable stimulation of the vagus nerve. With demonstrated potential in myriad applications, ranging from systemic inflammation reduction to traumatic stress attenuation, closed-loop tcVNS during periods of risk could improve treatment efficacy and reduce ineffective delivery. However, achieving this requires a deeper understanding of biomarker changes over time.

Objective: The aim of the present study was to reveal the dynamics of relevant cardiovascular biomarkers, extracted from wearable sensing modalities, in response to tcVNS.

Methods: Twenty-four human subjects were recruited for a randomized double-blind clinical trial, for whom electrocardiography and photoplethysmography were used to measure heart rate and photoplethysmogram amplitude responses to tcVNS, respectively. Modeling these responses in state-space, we (1) compared the biomarkers in terms of their predictability and active vs sham differentiation, (2) studied the latency between stimulation onset and measurable effects, and (3) visualized the true and model-simulated biomarker responses to tcVNS.

Results: The models accurately predicted future heart rate and photoplethysmogram amplitude values with root mean square errors of approximately one-fifth the standard deviations of the data. Moreover, (1) the photoplethysmogram amplitude showed superior predictability (P=.03) and active vs sham separation compared to heart rate; (2) a consistent delay of greater than 5 seconds was found between tcVNS onset and cardiovascular effects; and (3) dynamic characteristics differentiated responses to tcVNS from the sham stimulation.

Conclusions: This work furthers the state of the art by modeling pertinent biomarker responses to tcVNS. Through subsequent analysis, we discovered three key findings with implications related to (1) wearable sensing devices for bioelectronic medicine, (2) the dominant mechanism of action for tcVNS-induced effects on cardiovascular physiology, and (3) the existence of dynamic biomarker signatures that can be leveraged when titrating therapy in closed loop.

Trial Registration: ClinicalTrials.gov NCT02992899; https://clinicaltrials.gov/ct2/show/NCT02992899

International Registered Report Identifier (IRRID): RR2-10.1016/j.brs.2019.08.002

(JMIR Mhealth Uhealth 2020;8(9):e20488) doi:10.2196/20488

KEYWORDS

vagus nerve stimulation; noninvasive; wearable sensing; digital biomarkers; dynamic models; state space; biomarker; cardiovascular; neuromodulation; bioelectronic medicine

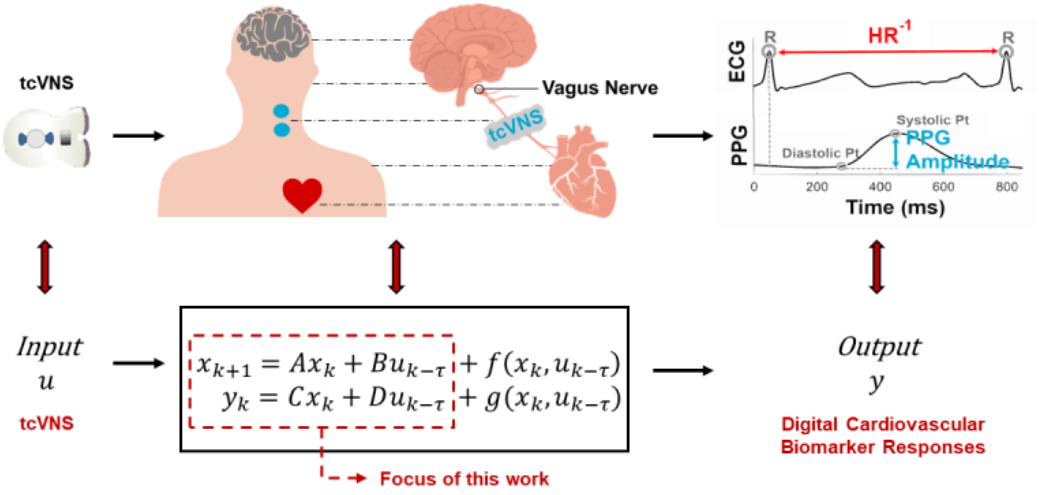
Introduction

Transcutaneous cervical vagus nerve stimulation (tcVNS) devices have emerged as inexpensive and convenient alternatives to implantable devices for stimulation of the cervical vagus nerve [1]. Over the last half decade, tcVNS-based devices have been approved by the Food and Drug Administration (FDA) for the treatment of migraine and cluster headache, and have demonstrated efficacy in myriad applications, ranging from pain and inflammation reduction to emotional/mental stress attenuation [2-5]. As a noninvasive, nonpharmacologic therapy with minimal side effects [6], the capacity for the widespread use of tcVNS is promising [7]. Yet, to achieve this potential in time-sensitive applications such as in response to traumatic stress, real-time information regarding the patient's state and response to tcVNS must be incorporated. Thus, a compelling need exists to understand observable biomarker dynamics in relation to tcVNS. This would advance scientific knowledge by uncovering temporal dependencies at finer time resolutions, facilitate real-time predictions of physiological responses to tcVNS for improved treatment titration, and pave the way for optimal estimation approaches to simultaneously track physiological state.

In enabling such closed-loop approaches to tcVNS, initial challenges include (1) identifying noninvasively measured biomarkers of desirable effects, (2) understanding how these biomarkers change dynamically with the delivery of tcVNS for a deepened understanding of the characteristic responses, and, accordingly, (3) predicting each subject's responses to tcVNS for improved treatment outcomes. In current psychiatric practice, the responses to interventions are primarily assessed through patient reports and qualitative judgments of symptoms [8]. However, these methods are limited by the subjectivity of such reporting and may not be reflective of the true nature of a psychiatric disturbance or subsequent therapeutic response due to factors that include disorder-based perceptual distortions (eg, dissociation), subjective bias, or alternative motivations (eg, malingering) [9]. Alternatively, by studying objective measures that reflect underlying physiological processes, these limitations could potentially be mitigated.

Unfortunately, the existing literature on such objective biomarkers is limited to invasive or obtrusive measures (eg, brain imaging or blood biomarkers) [4,10,11], next-day/longitudinal assessment [2], or static features extracted over minute-long time windows [5,12]. Improving upon this, we applied parametric modeling methods to digital (or physiological) biomarkers [13] that have been deemed to be the most promising in previous work [12]. Specifically, we used state-space models in an input-output formulation to model the dynamic responses of heart rate and photoplethysmogram (PPG) amplitude to tcVNS, as illustrated in Figure 1. State-space models were selected as the machine-learning framework of choice owing to their superior utility in real-time estimation and control applications [14,15]. This work thereby presents a first-of-its-kind dynamic analysis of physiological biomarker responses to tcVNS and furthers the state of the art by: (1) identifying PPG amplitude as a superior digital biomarker to heart rate for the prediction of real-time responses and appraisal of tcVNS effects; (2) quantifying a consistent delay in tcVNS-induced downstream cardiovascular biomarker variation; and (3) uncovering characteristic dynamic response signatures that separate PPG amplitude and heart rate responses to tcVNS from a sham stimulation.

Figure 1. High-level illustration of the modeling task investigated in this study. Transcutaneous cervical vagus nerve stimulation (tcVNS) is treated as the input to the underlying physiological system, while the observed cardiovascular responses found pertinent to tcVNS effects in previous work are treated as the output signals. This work focuses on the dynamics described by linear time-invariant difference equations, formulated as discrete-time state-space equations. The latent state is depicted here as variable x, while f and g are nonlinear functions. HR: heart rate; PPG: photoplethysmogram; ECG: electrocardiogram.



Methods

Human Subjects Experiment

As part of a collaborative study approved by the Institutional Review Boards of the Georgia Institute of Technology (#H17126), Emory University School of Medicine (#IRB00091171), SPAWAR Systems Center Pacific, and the Department of Navy Human Research Protection Program, 24 healthy human subjects, including 12 women and 12 men, with a mean age of 31 (SD 9) years, height of 173.4 (SD 8.9) cm, and weight of 77.9 (SD 13.7) kg, were recruited. All subjects had a history of prior psychological trauma, but without current posttraumatic stress disorder or other major psychiatric disorder, and written informed consent was obtained. Conforming to a randomized double-blind protocol spanning 3 days (ClinicalTrials.gov NCT02992899), each subject received four administrations—two on the first day and one on each of the following two days—of either "active" tcVNS or "sham" stimulation. These four administrations were accompanied by no other form of stimulus to focus solely on the effects of tcVNS on human physiology. Overall, 96 doses of either active or sham tcVNS were administered in total to the 24 subjects (with 12 allocated to the active group).

The active and sham devices (gammaCore, electroCore, Basking Ridge, NJ, USA) were identical in both appearance and operation, differing only with respect to the stimulation parameters. The active devices administered voltage signals consisting of five 5-kHz sinusoidal pulses repeating at a rate of 25 Hz, while the sham devices delivered biphasic square pulses at a rate of 0.2 Hz, resulting in a perceptible tingling sensation. For both active and sham devices, electrical stimulation was delivered transcutaneously to the left side of the neck, targeting the cervical vagus nerve projection. At specified times, the researcher initiated the device and adjusted the stimulation amplitude (ranging from 0 to 5 arbitrary units [AU], corresponding to 0-30 V and 0-14 V for the active and sham device, respectively) to as high as the subject could comfortably endure (active: 3.0 [SD 0.8] AU; sham: 4.4 [SD 1.2] AU). The amplitude was then kept fixed for the remainder of a 120-second period, following which the device automatically stopped, reducing its stimulation amplitude to zero. This 2-minute timeframe replicates the programmed administrations onboard the tcVNS devices that are currently in use [16]. For further protocol and stimulation details, the reader is referred to Gurel et al [5].

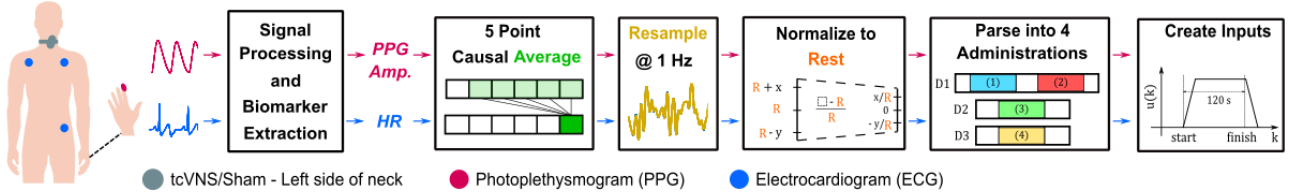
Physiological Sensing, Signal Processing, and Biomarker Extraction

Electrocardiogram (ECG) and transmissive PPG signals, taken from the finger, were continuously measured at the locations displayed in Figure 2 using the Biopac RSPEC-R and Biopac PPGED-R systems (Biopac Systems, Goleta, CA, USA), respectively. All data were acquired at a 2-kHz sampling rate using the Biopac MP150 16-bit data acquisition system.

To extract the instantaneous heart rate from the ECG signal, finite impulse response bandpass filtering (passband of 0.6-40 Hz) was applied to cancel out-of-band noise while maintaining the waveform shape [17]. By then detecting the R-peaks, the instantaneous heart rate was computed in beats per minute by taking the reciprocal of the time length, in minutes, between each pair of R-peaks (R-R interval). These R-peaks were subsequently leveraged to beat-separate the bandpass-filtered PPG signals (passband of 0.4-8 Hz [18]). PPG amplitude (in AU), was then calculated on a beat-by-beat basis by subtracting the global minimum of each PPG beat from its global maximum.

The focus on heart rate and PPG amplitude as the biomarkers of interest for this work was based on a rationale that both cardiac and vascular downstream responses to stimulus were of interest scientifically and may indicate different autonomically mediated mechanisms following tcVNS. Heart rate is a hallmark measure of the cardiac response to changes in autonomic tone and is controlled by both the sympathetic and parasympathetic branches of the autonomic nervous system. Parasympathetic decreases in heart rate are mediated by the release of acetylcholine that binds to muscarinic receptors in the heart, whereas sympathetic increases in heart rate are mediated by the release of epinephrine and norepinephrine that bind to beta-1 receptors in the heart. PPG is a measure of peripheral blood volume pulse, and represents a surrogate measure of vasodilation or vasoconstriction resulting primarily from changes in sympathetic tone. Peripheral vasoconstriction is mediated by alpha-1 receptors in the smooth muscle of the vasculature [19].

Figure 2. Physiological sensing, signal processing, and modeling preparation. Twenty-four subjects (12 active) underwent a clinical protocol, wherein the electrocardiogram (ECG) was measured with electrodes placed in a three-lead configuration and the photoplethysmogram (PPG) was measured from the fingertip in a transmissive LED-photodiode setup. Transcutaneous cervical vagus nerve stimulation (tcVNS) or sham stimulation was administered at predefined times, where the exact stimulation location was identified by locating the left carotid pulse. Following signal processing and biomarker extraction, the biomarkers were prepared for modeling via 5-point causal averaging, resampling to 1 Hz, normalizing to rest, and finally parsing into 4 separate vectors associated with the 4 tcVNS/sham administrations studied. By referencing stimulation initiation, the corresponding input amplitude waveforms were then constructed to replicate device administration. Amp.: amplitude; D1: day 1; D2: day 2; D3: day 3.



Preparing the Biomarker Time Series for Modeling

Following feature extraction, the heart rate and PPG amplitude values existed as beat-to-beat time series of nonuniform sampling rates (due to variability in heart rate). Thus, prior to any modeling, a few time-series processing steps were employed (Figure 2). First, each subject's biomarker time series were prepared using a causal moving average of 5 data points (approximately 5-second, rectangular windows) to attenuate high-frequency artifacts. The filtered time series were then resampled at a constant frequency of 1 Hz to satisfy uniform sampling rate requirements [20], followed by normalization to each day's resting value (subtracting and then dividing by rest) to account for intersubject variability during our subsequent population-level analysis.

To then separately investigate the effects of tcVNS/sham administration on each of the two biomarkers, the resultant time series were parsed according to the 4 administrations per subject. Based on the data available, parsing was achieved by leaving 60 seconds of data prior to each 120-second administration and retaining 120 seconds of data postadministration, for a total of 300 seconds per administration. Note that missing data at the end of the 300-second interval relevant to our analysis affected 3 administrations among the 96 collected, and therefore the corresponding data vectors were shortened accordingly.

The accompanying input data were then created for each of the 192 subject-administration-biomarker combinations (2 biomarkers × 4 administrations × 24 subjects). This was accomplished by modeling the relative tcVNS/sham amplitude delivered to each subject with pulse-like trapezoidal signals that replicated the ramping and stabilization of true stimulation. These inputs were formed by passing a boxcar input of 120-second width and unit amplitude through a 5-point moving average filter. Since stimulation amplitude remains the only variable modulated during tcVNS/sham administration, stimulation amplitude was specified as the input variable, as done in related work [21,22]. This ensured that the subsequently analyzed input-induced effects modeled the tcVNS-induced effects under study. Note that the digital biomarker response dynamics modeled in this study correspond to a particular therapy—FDA-approved tcVNS—that exhibits equivalent input bandwidths across all administrations. Therefore, modeling the input-output relationship for the specific input variability exhibited during practical device usage remains invaluable to future analysis and development. For further reasoning and evidence behind this approach, please see Multimedia Appendix 1.

State-Space Modeling and Cross-Validation

Considering the established diversity in VNS outcomes, which itself remains an active area of research [23], subject-specific models were developed for our study, leaving the identification of consistencies across relevant subject groupings for the subsequent model analysis phase. To address biomarker differences, heart rate and PPG amplitude were modeled separately to foster comparison between the estimated systems and responses for each of the two biomarkers. This maneuver to disentangle outputs that seemingly respond to tcVNS/sham simultaneously derives from a result in state-space input-output modeling, where the absence of output feedback–controlled effects on the input under study allows for the disentangling of multiple-output systems [24] (see Multimedia Appendix 1).

The forthcoming equations and corresponding explanations will thus depict single-input single-output systems. The discrete-time state-space model structure in innovations form is governed by the following two difference equations:

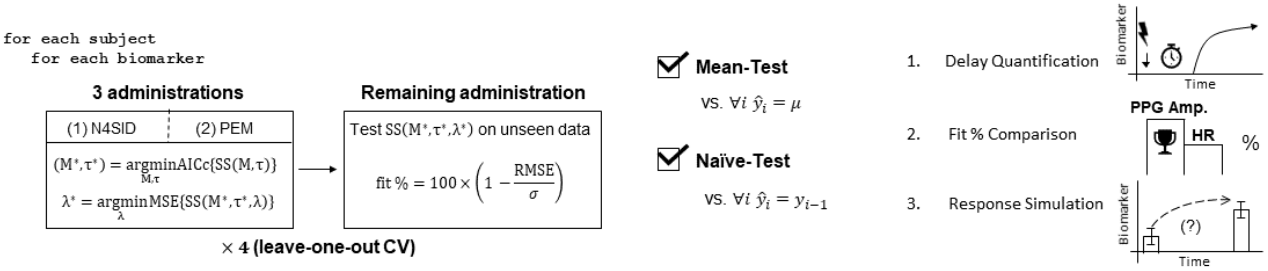
$$x_{k+1}=Ax_k + Bu_{k-\tau} + Ke_k \quad \textbf{(1)}$$

$$y_k=Cx_k + Du_{k-\tau} + e_k \quad \textbf{(2)}$$

where k represents the discrete time step, τ represents the input delay of the system (also known as the "dead time" between changes in input and resultant changes in system behavior), $y \in \mathbf{R}$ represents the output, $u \in \mathbf{R}$ represents the input, $x \in \mathbf{R}^M$ represents the latent state of the dynamical system (referred to as the state vector), where M is referred to as the model order of the system, $e \in \mathbf{R}$ is the residual computed as the innovation estimate, and $A \in \mathbf{R}^{M\times M}$, $B \in \mathbf{R}^{M\times 1}$, $C \in \mathbf{R}^{1\times M}$, $D \in \mathbf{R}^{M\times M}$, and $K \in \mathbf{R}^{M\times 1}$ are matrices consisting of free parameters that are estimated to best describe the data provided. In the context of this work, the scalar D is set to zero, as it quantifies the feedthrough component's contribution on the output (ie, the static contribution of the input that circumvents dynamics entirely).

In this study, we initialized model estimates using subspace identification [25] and then employed prediction error minimization to ameliorate any limitations of the computationally cheaper subspace method [26]. In accordance with our subject-specific modeling objective, leave-one-out cross-validation was used to train and evaluate the state-space models. Specifically, each of the 4 administrations was used for testing exactly once, with the remaining 3 administrations used for model estimation. This eventually resulted in an overall 4 models per biomarker per subject, following the optimization process detailed below. Figure 3 illustrates this cross-validation process.

Figure 3. Modeling, optimization, and cross-validation; quality assurance; and analysis. (Left) Modeling, optimization, and testing were performed in a leave-one-out cross-validation (CV) scheme, where, out of the four administrations, each administration was considered once as the unseen test set. The model order (M) and input delay (τ) were optimized by minimizing the small sample size-corrected Akaike information criterion (AICc). For each model order-input delay combination, a state-space model was trained by first initializing the parameter estimates using subspace estimation (N4SID); this was then followed by prediction error minimization (PEM) to refine the parameter estimates. Ridge regression was then performed for this specific model configuration by iterating lambda (λ) logarithmically over a specified interval and minimizing the mean square error. The 1-step-ahead prediction performance was evaluated using a fit percentage formula based on the root mean square error (RMSE) normalized by the standard deviation of the data (σ). (Middle) For quality assurance beyond subjective satisfaction in visual results, the models were evaluated against two objective baseline tests from the literature: the mean test and the naïve test. (Right) To extract the model information pertinent to a deepened dynamic understanding of biomarker responses, (1) optimal input delays were compiled to assess the expected response latency following stimulation onset, (2) the biomarkers were compared against each other to identify superiority in monitoring dynamic changes following transcutaneous cervical vagus nerve stimulation (tcVNS), and (3) the responses were quantifiably visualized via controlled simulation to produce population trajectories of expected dynamic changes following transcutaneous cervical vagus nerve stimulation (tcVNS) administration, in comparison to sham.



Regularization and Hyperparameter Optimization

The small sample size-corrected Akaike information criterion (AICc) was employed to select the optimal model configuration (M*, τ*); the AICc was used instead of the standard AIC due to the ratio between training data points, *N*, and the number of parameters, *p*, in the largest candidate model remaining less than 40 [27,28]. For technical details on usage, please refer to Multimedia Appendix 1.

To optimize τ, based on a previous effort to subjectively annotate the delays between tcVNS and biomarker changes [29], the interval was set to $0 \leq \tau \leq 35$; this was to circumscribe a 99% confidence interval about the reported result of 18 seconds (SD 7). For *M*, the lack of broadband input constrained the parameter total to below the order of input persistence of excitation for any candidate model [30]. The order of persistent excitation is estimated by counting the number of distinct frequencies with spectral content larger than a set threshold [31]. By performing this computation iteratively across all input signals used in this study, including those associated with datasets missing data points, the order never decreased below 50. Thus, our optimization interval was safely restricted to a maximum of 50 parameters, corresponding to model order 10 for modal form estimates (see Multimedia Appendix 1).

Once the optimal state-space model was selected for each training set of 3 administrations, the model was then regularized separately via ridge regression [32]. The parameter λ was selected by minimizing the mean square error over the interval $\lambda \in [10^{-15}, 10^{4}]$. The lower bound was selected due to MATLAB's machine epsilon for double-precision floating points (2^{-52}) and the upper bound was selected during an experimentation period. These optimal, regularized model estimates were then used for the remaining analyses. The methods described in this subsection and the previous section correspond to the box on the left in Figure 3.

Fit Percentage Definition and Baseline Testing

The models were evaluated against two established baselines for dynamic modeling tasks: the mean test and the naïve test. The mean test involves comparing out-of-sample 1-step prediction performance vs mean predictors, and the naïve test involves comparing 1-step prediction performance vs the naïve predictor [33-35] (see Multimedia Appendix 1 for further details). This out-of-sample testing is depicted in the box on the right of Figure 3 (left). Figure 3 (middle) also summarizes the baseline testing.

The metric used herein for evaluation is the fit %, defined as:

where $\hat{y}=[\hat{y}_1\hat{y}_2\ldots\ \hat{y}_N]^T$ represents the predicted output values from time step 1 to N and $y=[y_1y_2\ldots\ y_N]^T$ represents the true output. This exact metric has been widely used to quantify time-series model validity (eg, [21,24]), along with its variants (eg, [33]). Note that the above fit % formula can be equivalently rewritten as $(1 - RMSE/\sigma) \times 100\%$, where RMSE is the root mean square error between the predicted and true values and σ is the standard deviation of the data. Thus, the fit % used here is commonly referred to as a fit % metric based on standard deviation – normalized RMSE, which represents an estimate of the output variability the model can accurately reproduce [24]. This is seen to be the case when comparing the fit % to the coefficient of determination, given by:

demonstrating that the fit % simply replaces the mean square error and variance with the RMSE and standard deviation, respectively. It thereby exhibits improved spread for $RMSE < \sigma/2$.

Model Configuration and Fit Comparisons

Figure 3 (right) illustrates the subsequent methods of analysis. To determine the presence of any notable differences in

predictability between the biomarkers in question, the biomarker fit percentages were compared. Additionally, the optimal model orders associated with each biomarker's models were compared to determine whether any model complexity differences could be posited. In this case, unlike fit %, lower model orders are generally favored, as they signify relatively simpler systems.

To further investigate the optimal model configurations, the automatically optimized input delays, τ^*, for each of the two biomarkers were compiled. These two sets of latencies were then compared against the manually annotated delays of previous work. As detailed in Gurel et al [29], this manual labeling was performed independently by three investigators with an interannotator agreement of 90%.

Statistical Testing

With 4 fit % values, model orders, and input delays obtained per subject-biomarker combination (corresponding to the 4 models produced via cross-validation), all quantities were first averaged across all 4 models prior to comparison/compilation. To examine biomarker differences in fit % and model order, pairwise *t* tests or Wilcoxon signed-rank tests were employed for normally and nonnormally (tested using the Shapiro-Wilk test) distributed variables, respectively. For comparison of the delay results, after failing to reject normality and sphericity (tested using the Mauchly sphericity test), a one-way repeated-measures analysis of variance was used; α=.05 was used as the level of significance for these comparisons.

Investigating Biomarker Responses to tcVNS vs Sham

As a final analysis step, we investigated the tcVNS-induced response dynamics captured through modeling by simulating both the active and sham models forward from a controlled state. To leverage the previously resampled biomarker time series, we assembled a second set of plots corresponding to the true experimental responses. To facilitate qualitative comparisons, both sets were constructed by compiling/simulating the true/artificial biomarker time series such that 10 seconds existed prior to the true/simulated stimulus administration and 120 seconds remained afterward. The 120 seconds corresponds to the 2-minute poststimulus period used during modeling, and the 10 seconds were included to better understand the true biomarker values prior to stimulation.

Resampled Experimental Responses

For each of the 4 administrations for the 24 subjects, the heart rate and PPG amplitude time series were extracted by simply considering the values produced as a result of the modeling preparation steps. With resampled and normalized time series in hand, each biomarker's overall response was constructed by first averaging the biomarker responses across all 4 administrations, followed by additional average and SEM calculations across all 12 subjects in each device group.

Simulated Model Responses

To visualize the biomarker dynamics captured by the models, each model was simulated forward by (i) setting the initial conditions to zero (ie, initializing the system at its equilibrium) to guarantee equivalence between active and sham time series (setting x_1=0 guarantees that the system will remain there for as long as no stimulus is present); (ii) constructing an input waveform equivalent to the input used during the modeling process, nonzero between time points 10 and 130 seconds; and (iii) solving the difference equations forward in time for each model, ignoring the contribution of *Ke* and *e* in Equations (1) and (2), respectively, as these quantities represent the unmodeled aspect of the system, along with noise, lumped into residual terms [36]. Each biomarker's overall response was constructed by averaging across all 4 simulated model responses, followed by calculation of the average and SEM across each device group. For further details regarding the advantages of this approach and insight into the subsequent simulation analysis, please refer to Multimedia Appendix 1.

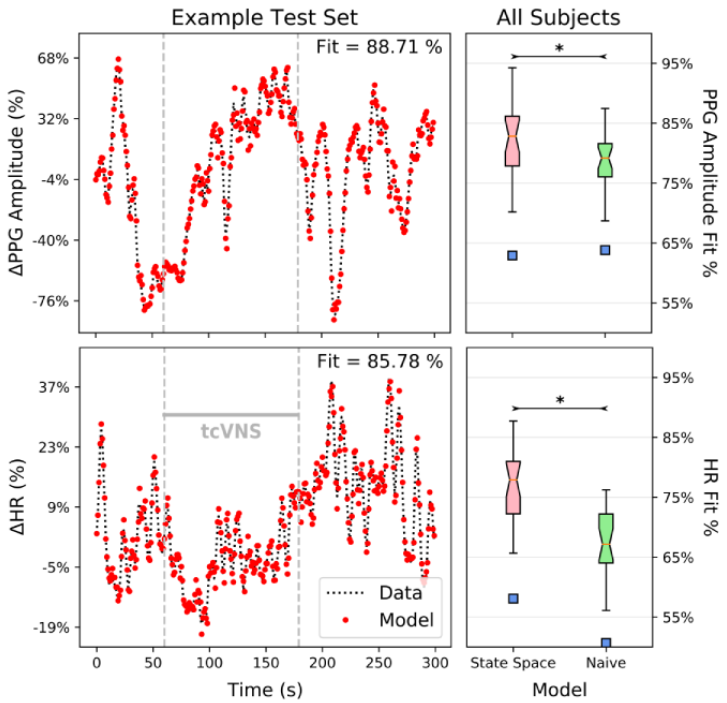
Results

Baseline Test Results

Figure 4 (left) displays the heart rate and PPG amplitude data in response to a single administration of active tcVNS from a representative subject. For illustrative purposes, the corresponding model predictions are overlaid. To produce these predictions for each biomarker, the model was trained on the remaining three datasets for this subject and tested on this particular dataset.

The naïve test results across all subjects are shown on the right side of Figure 4. Note that the mean predictors, by definition, always produce fits of 0%. Hence, the models for both biomarkers passed both the naïve test and the mean test, exhibiting significantly higher fit percentages than the naïve predictors (P<.001) and mean predictors (P<.001).

Figure 4. Example model predictions vs true data, along with naïve test box plot results for the entire sample. (Left) Both model vs data plots shown originate from one administration of a single active transcutaneous cervical vagus nerve stimulation (tcVNS) subject. Each biomarker's model was trained on the three other tcVNS administration datasets available for this subject; the test results on the remaining unseen dataset in a 1-step-ahead prediction task are shown. The gray dashed lines demarcate the time frame in which tcVNS was administered, and the fit percentages are calculated as previously defined. The y-axis represents relative changes from rest in percent form. (Right) The regularized, optimal state space models strongly satisfied both the naïve test and the mean test. The statistically significant (P<.001) naïve test results are shown, where the box plots indicate results for the entire sample; * denotes statistical significance. The blue squares indicate outlier points, where outliers lie above the upper quartile or below the lower quartile by more than 1.5 times the interquartile range. Significance (P<.001) held for the mean test as well for both biomarkers (not shown).



Modeling Amenability Comparison

Table 1 summarizes the results from the biomarker fit % and model order contrast. The fit of each of the PPG amplitude models was significantly better than that of the heart rate models (P=.03), albeit without compensation through an increase in model order/complexity. Model order did not differ between heart rate and PPG (P=.14).

Table 1. Mean (SD) fit % and model orders for each biomarker's state-space model.

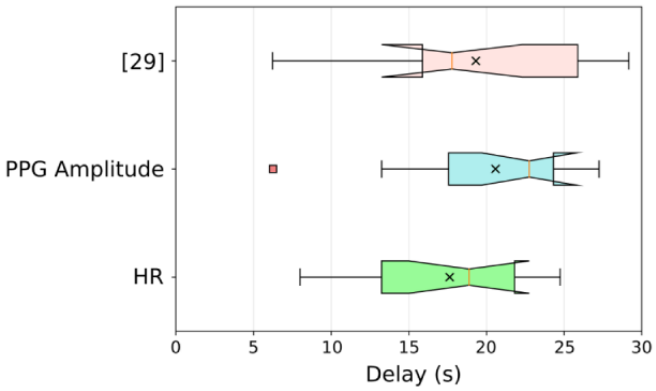
Metric	Heart rate	PPG[a] amplitude
Fit %	76.67 (6.96)	81.64 (7.07)
Model order	9.52 (0.38)	9.39 (0.34)

[a]PPG: photoplethysmogram.

Automatically Estimated Input Delays vs Manually Labeled Onset Delays

Figure 5 shows the automatically estimated input delays for both biomarker models, as well as the manually annotated result from previous work [29]. The mean input delay for the heart rate and PPG amplitude state-space models was 17.65 seconds (SD 5.17) and 20.58 seconds (SD 5.81), respectively, and the mean manually labeled delay was 18 seconds (SD 7). No statistically significant differences were found between the three sets of onset delays ($F_{2, 22}$=0.71, P=.51, η^2_p=.06).

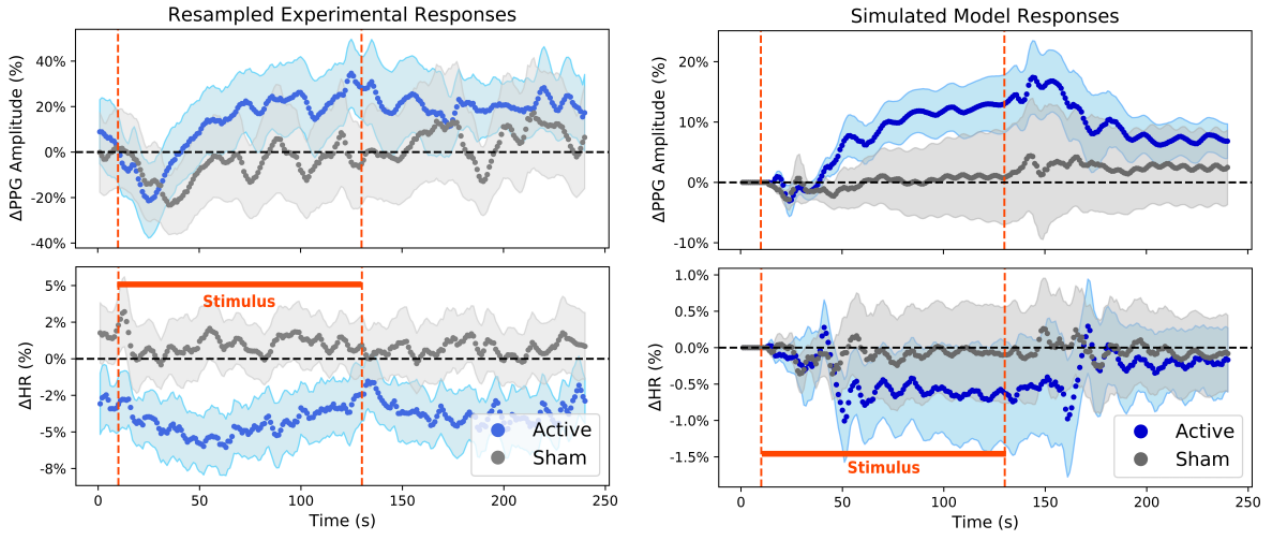
Figure 5. Delayed effects of transcutaneous cervical vagus nerve stimulation (tcVNS) in downstream cardiovascular biomarkers. By including input delay as a free parameter necessitating optimization, the objective, quantitative state-space model optimization process described in this paper reproduced the tcVNS delay findings of the manual, subjective annotations of [29]; no statistically significant differences existed between the three sets of onset delays. The coral-colored square represents an outlier, and the black crosses represent the means.



Biomarker Responses to tcVNS vs Sham

Figure 6 shows the heart rate and PPG amplitude responses to tcVNS or sham administration. The graphs on the left correspond to the resampled experimental responses, and those on the right correspond to the simulated model responses. Although conclusions are weakened when comparing the true active and sham responses due to the lack of controlled initial conditions, one can posit possible dynamic response signatures for PPG amplitude and heart rate in response to active tcVNS when comparing against their respective sham counterparts. Through modeling and simulation, we remedy these initial condition concerns. Rubin causality is therefore exhibited in response to tcVNS, allowing for causal inference to be legitimately made by comparing the active and sham responses. Moreover, the underlying dynamic response signatures of tcVNS can be better visualized in comparison to the uneventful sham responses.

Figure 6. Dynamic responses to transcutaneous cervical vagus nerve stimulation (tcVNS) for heart rate (HR) and photoplethysmogram (PPG) amplitude. The curves themselves, along with their accompanying shaded regions, represent average (SEM). (Left) True responses, resampled for population averaging and subsequent dynamic modeling. The y-axis values represent relative changes from rest in percent form, and the orange dashed lines demarcate the period of active/sham stimulus administration ($t \in (10, 130)$ seconds). (Right) Simulated responses to tcVNS, produced by solving the state-space model difference equations forward in time. The y-axis values represent relative changes from rest in percent form, and the orange dashed lines demarcate the period of simulated active/sham stimulus administration ($t \in (10, 130)$ seconds).



Discussion

PPG Amplitude is a More Reliable and Predictable Biomarker of tcVNS Effects

Based on the results in Table 1, we found that the models were able to reproduce the PPG amplitude data significantly better than the heart rate data, albeit without requiring any increase in model complexity to compensate. PPG amplitude may thus be a more trackable dynamic marker of the physiological responsiveness to tcVNS than those previously investigated. Notably, such biomarkers have thus far remained elusive in the existing tVNS literature [37]. Hence, in conjunction with the consistency of PPG amplitude discovered in previous work [5,12], the finding that PPG amplitude remains more amenable to our modeling approach suggests that PPG amplitude may serve as a superior biomarker for real-time tcVNS systems, and thus should be given precedence in a multimodal sensing setting.

Biomarker characteristics that could explain this distinction involve measurement location and physiological origin. Heart rate is a cardiac measure originating centrally from the heart, whereas the PPG signals measured in this study were sensed at the periphery—in particular, transmissively at the fingertip. As described in the Methods section, a clear physiological distinction exists; these differences may in fact explain the relative contrasts observed in this study, as well as the previous difficulties encountered in identifying a trustworthy biomarker derived from the heart (eg, heart rate variability measures) [37].

Cardiovascular biomarkers extracted from peripheral processes, mediated solely by the sympathetic nervous system (eg, vasoconstrictor sympathetic nerve activity [38] and PPG amplitude [18]), have shown repeated success in capturing the effects of tVNS. An explanation for this may involve physiological control. In comparison to the peripheral blood vessels and their sympathetically mediated vasoconstriction, heart rate is regulated by both the sympathetic and parasympathetic nervous systems; additionally, the heart, owing to its critical role in maintaining blood supply to the entire body, is subject to far tighter regulation [19,39]. The peripheral vasculature thus remains more susceptible to external modulation, or in a systems sense, external "disturbances." Thus, a promising approach for sense-and-react systems that seek to assess the effects of external modulation may be to leverage peripheral biomarkers that measure quantities subject to looser homeostatic control.

tcVNS-Induced Effects are Delayed in Digital Cardiovascular Biomarkers

Figure 5 illustrates the consistency found between the optimized input delays of this work and the manually annotated latencies of prior work [29]. Note, however, that these delay values of approximately 15 seconds may in fact be overestimates in certain situations due to the approximate 5-second ramp-up period during administration. Since this aspect was accounted for during modeling, one can reasonably deduct 5 seconds from the delays listed for purposes involving closed-loop system design and hypothesizing the dominant mechanism of action. Our reasoning is that for a closed-loop system adapted for a particular subject, the threshold of comfort will likely be known, thereby eliminating the ramp-up period. To provide quantitative evidence for a dominant mechanism of action, protocol-related delays are not necessarily relevant to determining the underlying pathways that tcVNS affects when causing downstream cardiovascular effects. Nevertheless, as determined by the notched confidence intervals shown in Figure 5, mechanistically relevant delays remain at latencies of >5 seconds, even after factoring in the 5-second deduction. Thus, for closed-loop tcVNS systems, a delay greater than 5 seconds must be taken into consideration and designed for accordingly [40,41].

In providing quantitative evidence for the likely dominant mechanism of action for tcVNS effects on downstream physiology, we here highlight the two prevalent hypotheses for tcVNS at either the cervical or auricular branch of the vagus nerve [42,43] (see Multimedia Appendix 1 for a detailed illustration). The first hypothesis involves electrical activation of vagal efferents terminating at the heart [44]; this mechanism would in fact induce the expected decrease in heart rate, which has been frequently cited in previous animal studies (eg, [45]) and was also found to be the case here. However, this hypothesis may not necessarily explain the sympathetically mediated effects observed at the periphery during tVNS administration (eg, vasoconstrictor sympathetic nerve activity [38]). The second hypothesis implicates afferent vagal stimulation as a pathway to brain activity in autonomically relevant brain areas, followed by downstream effects induced by resultant efferent signaling. Although brain imaging and electrophysiological measurements have confirmed the activation of vagal afferents during tcVNS [10,46], it remains unclear whether these "bottom-up" effects serve as the dominant cause for the resultant autonomic efferent activity.

The results presented herein seem to align with the latter hypothesis: namely, that the delayed effects modeled—and subjectively observed in previous work—are a byproduct of afferent vagal activity, processing in the brain, and resultant efferent-mediated autonomic effects. Synthesizing previous sensing and measurement studies of tcVNS, afferent signatures (P1-N1 vagal somatosensory evoked potentials) tend to occur within 1 second of amplitude stabilization [46]. Considering that finite element modeling results suggest that only A and B vagal fibers can be stimulated through gammaCore tcVNS [47], and that efferent B fiber conduction velocities range from 5 to 10 m/s [44], we would expect that if the dominant pathway for cardiac effects occurred through direct efferent stimulation, these effects would occur within 2-3 seconds of amplitude stabilization. Instead, we conclude that the dominant tcVNS effects on cardiovascular physiology likely occur after 5-10 seconds have elapsed, following amplitude stabilization (after reasonably deducting a 5-second ramp-up period from the ~15-second delays reported). This suggests that downstream cardiovascular effects are likely mediated by the prolonged afferent-brain-efferent mechanism of action.

Distinct Dynamic Signatures Characterize tcVNS-Induced Effects

Heart rate and PPG amplitude responses were visualized both for resampled data averaged across administrations and subjects, and for simulated responses produced by solving the estimated difference equations forward in time. As shown in Figure 6 (left), the average active PPG amplitude response to tcVNS clearly exceeded the average sham response to tcVNS, although they coincidentally initialized at similar relative values. In contrast, the average active heart rate graph displays a transient decrease in response to tcVNS that lasts about 50 seconds, followed by a return to prestimulus values. Interestingly, this agrees with recent findings in the auricular tVNS literature, which noted 3-4% transient drops in heart rate that recovered over the course of 30 seconds following stimulation onset [48]. Unlike the PPG amplitude responses, the average heart rate responses of the two device groups initialized with an almost 4% relative difference between them, which is an important point to consider when comparing the two trajectories. Although the average active heart rate response seems starkly different from the average sham heart rate response, if one were to artificially shift the two responses' initial conditions vertically to level the playing field, the active response would remain

beneath the sham response, but definitively by less. Moreover, this disregards the concerns associated with performing such an artificial transformation for causal inference and comparison [49].

This issue of Rubin causality (see Multimedia Appendix 1) is ultimately resolved on the right side of Figure 6. Since the dynamic models developed thus far were simulated forward from the same equilibrium condition, we arrived at population-level characteristic responses to tcVNS that exhibited the necessary equivalent prestimulus behavior. With this added causal inference power, we not only observed apparent similarities between the plots on the left and right of Figure 6, but we also discovered evident differences between the active and sham groups in their responses to tcVNS. In particular, the active group's simulated PPG amplitude responses significantly exceeded the relatively flat sham PPG amplitude trajectories; analogously, but in the opposite direction, the active heart rate trajectories displayed a clear decrease in comparison to the sham responses, where the sham responses again remained relatively constant. Thus, the exact same modeling and simulation methodology were applied to data from both device groups, and yet stark differences arose in the characteristic responses to stimulation. Furthermore, the sham group displayed relatively subdued responses to the simulated stimulus, as expected. These results serve as further validation for the current approach and ultimately imply the presence of distinctive dynamic signatures that characterize the continuous-time effects of tcVNS on digital cardiovascular biomarkers.

Limitations and Future Work

A few limitations are to be noted for this study. Although nonlinear approaches to difference equation modeling are generally discouraged at the outset unless expert knowledge or sufficient evidence seems to suggest otherwise [24], nonlinear dynamics exist in general. Nevertheless, in agreement with previous findings demonstrating that a considerable percentage of dynamic variability exhibited by cardiovascular biomarkers such as heart rate can be modeled linearly [50], we found that approximately 80% of the biomarker variability observed can still be predicted accurately. Hence, this paper presents a "best linear approximation" that could further be built upon in future work directed at characterizing the nonlinearities of tcVNS responses [24].

Although all tcVNS clinical protocols reported to date have used the gammaCore device to which the present results readily apply, future control approaches will need to venture beyond this standardized waveform and vary parameters during stimulation to achieve desired regulation goals, while simultaneously maintaining user safety. This would additionally help in satisfying the stringent input requirements for system identification [24], which are conditions that have not been met thus far.

Finally, we note that biomarkers other than PPG amplitude and heart rate were found to be statically relevant in quantifying tcVNS effects in previous work [5], although static PPG amplitude and heart features were found to be the most salient [12]. Thus, a multimodal closed-loop system that utilizes signals other than ECG and PPG may benefit from further application of such dynamic modeling and analysis.

Conclusions

In this work, we studied heart rate and PPG amplitude responses to tcVNS and derived three key findings by approaching this question from a dynamic modeling perspective. First, PPG amplitude demonstrates preeminence in both modeling amenability and active vs sham response separation, suggesting its superiority as a digital biomarker for real-time response prediction and tcVNS effect quantification. Second, a consistent delay of greater than 5 seconds exists between tcVNS onset and downstream cardiovascular biomarker responses. Latency must therefore be considered and accounted for appropriately during clinical monitoring and closed-loop system design. Moreover, this delay may provide measurable evidence for the dominance of the hypothesized vagal afferent-to-brain-to-autonomic efferent pathway in downstream cardiovascular modulation. Lastly, state-space models can successfully predict heart rate and PPG amplitude dynamics in response to tcVNS, and can help to identify the characteristic dynamic signatures that separate these biomarker responses to tcVNS from sham stimulation. This dynamic modeling and analysis thereby deepens our understanding of tcVNS effects and lays the groundwork for future closed-loop approaches in time-sensitive applications.

Acknowledgments

This work was sponsored in part by the Defense Advanced Research Projects Agency (DARPA) Biological Technologies Office (BTO) Targeted Neuroplasticity Training (TNT) program through the Naval Information Warfare Center (NIWC) Cooperative Agreement No. N66001-16-4054, and by DARPA BTO ElectRx through NIWC Cooperative Agreement No. N66001-19-2-4002. AS was funded by the National Institutes of Health (K23HL127251). JB receives funding from an electroCore Investigator Initiated Research Grant and from a Distinguished Investigator Award from the Brain and Behavior Foundation. The gammaCore stimulation devices used in this study were provided by electroCore.

Conflicts of Interest

JB has received funding from an electroCore Investigator Initiated Research Grant, and the gammaCore stimulation devices used in this study were provided by electroCore.

Multimedia Appendix 1
Supplementary material.

[PDF File (Adobe PDF File), 330 KB - mhealth_v8i9e20488_app1.pdf]

References

1. Adair D, Truong D, Esmaeilpour Z, Gebodh N, Borges H, Ho L, et al. Electrical stimulation of cranial nerves in cognition and disease. Brain Stimul 2020 May;13(3):717-750 [FREE Full text] [doi: 10.1016/j.brs.2020.02.019] [Medline: 32289703]
2. Brock C, Brock B, Aziz Q, Møller HJ, Pfeiffer Jensen M, Drewes AM, et al. Transcutaneous cervical vagal nerve stimulation modulates cardiac vagal tone and tumor necrosis factor-alpha. Neurogastroenterol Motil 2017 May 12;29(5):e12999. [doi: 10.1111/nmo.12999] [Medline: 27957782]
3. Lerman I, Davis B, Huang M, Huang C, Sorkin L, Proudfoot J, et al. Noninvasive vagus nerve stimulation alters neural response and physiological autonomic tone to noxious thermal challenge. PLoS One 2019 Feb 13;14(2):e0201212 [FREE Full text] [doi: 10.1371/journal.pone.0201212] [Medline: 30759089]
4. Lerman I, Hauger R, Sorkin L, Proudfoot J, Davis B, Huang A, et al. Noninvasive Transcutaneous Vagus Nerve Stimulation Decreases Whole Blood Culture-Derived Cytokines and Chemokines: A Randomized, Blinded, Healthy Control Pilot Trial. Neuromodulation 2016 Apr 15;19(3):283-290. [doi: 10.1111/ner.12398] [Medline: 26990318]
5. Gurel NZ, Huang M, Wittbrodt MT, Jung H, Ladd SL, Shandhi MMH, et al. Quantifying acute physiological biomarkers of transcutaneous cervical vagal nerve stimulation in the context of psychological stress. Brain Stimul 2020 Jan;13(1):47-59 [FREE Full text] [doi: 10.1016/j.brs.2019.08.002] [Medline: 31439323]
6. Redgrave J, Day D, Leung H, Laud P, Ali A, Lindert R, et al. Safety and tolerability of Transcutaneous Vagus Nerve stimulation in humans; a systematic review. Brain Stimul 2018 Nov;11(6):1225-1238. [doi: 10.1016/j.brs.2018.08.010] [Medline: 30217648]
7. Bremner JD, Rapaport MH. Vagus Nerve Stimulation: Back to the Future. Am J Psychiatry 2017 Jul 01;174(7):609-610. [doi: 10.1176/appi.ajp.2017.17040422] [Medline: 28669203]
8. Venigalla H, Mekala HM, Hassan M, Ahmed R, Zain H, Dar S, et al. An Update on Biomarkers in Psychiatric Disorders - Are we aware, Do we use in our clinical practice? Ment Health Fam Med 2017;13:471-479.
9. Pitman RK, Rasmusson AM, Koenen KC, Shin LM, Orr SP, Gilbertson MW, et al. Biological studies of post-traumatic stress disorder. Nat Rev Neurosci 2012 Nov 10;13(11):769-787 [FREE Full text] [doi: 10.1038/nrn3339] [Medline: 23047775]
10. Frangos E, Komisaruk BR. Access to Vagal Projections via Cutaneous Electrical Stimulation of the Neck: fMRI Evidence in Healthy Humans. Brain Stimul 2017 Jan;10(1):19-27. [doi: 10.1016/j.brs.2016.10.008] [Medline: 28104084]
11. Wittbrodt MT, Gurel NZ, Nye JA, Ladd S, Shandhi MMH, Huang M, et al. Non-invasive vagal nerve stimulation decreases brain activity during trauma scripts. Brain Stimul 2020 Jul 10;13(5):1333-1348 [FREE Full text] [doi: 10.1016/j.brs.2020.07.002] [Medline: 32659483]
12. Gurel NZ, Wittbrodt MT, Jung H, Ladd SL, Shah AJ, Vaccarino V, et al. Automatic Detection of Target Engagement in Transcutaneous Cervical Vagal Nerve Stimulation for Traumatic Stress Triggers. IEEE J Biomed Health Inform 2020 Jul;24(7):1917-1925. [doi: 10.1109/jbhi.2020.2981116]
13. Coravos A, Khozin S, Mandl KD. Erratum: Author Correction: Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019 May 10;2(1):40 [FREE Full text] [doi: 10.1038/s41746-019-0119-8] [Medline: 31304386]
14. Kailath T, Sayed AH, Hassibi B. Linear Estimation. Upper Saddle River, NJ: Prentice-Hall; 2000.
15. Ioannou P, Sun J. Robust adaptive control. Mineola, NY: Dover Publications; 2012.
16. Instructions for Use for gammaCore Sapphire. 2019. URL: https://www.gammacore.com/wp-content/themes/gammacore-p2/pdf/gammacore-IFU.pdf [accessed 2020-05-20]
17. Sornmo L, Laguna P. Bioelectrical Signal Processing in Cardiac and Neurological Applications. Burlington, MA: Elsevier Academic Press; 2005.
18. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Feb 20;28(3):R1-R39. [doi: 10.1088/0967-3334/28/3/r01]
19. Klabunde RE. Cardiovascular Physiology Concepts. Baltimore, MD: Lippincott Williams & Wilkins; 2012.
20. Xiao X, Mullen TJ, Mukkamala R. System identification: a multi-signal approach for probing neural cardiovascular regulation. Physiol Meas 2005 Feb 02;26(3):R41-R71. [doi: 10.1088/0967-3334/26/3/r01]
21. Santaniello S, Fiengo G, Glielmo L, Grill WM. Closed-Loop Control of Deep Brain Stimulation: A Simulation Study. IEEE Trans Neural Syst Rehabil Eng 2011 Feb;19(1):15-24. [doi: 10.1109/tnsre.2010.2081377]
22. Yang Y, Connolly AT, Shanechi MM. A control-theoretic system identification framework and a real-time closed-loop clinical simulation testbed for electrical brain stimulation. J Neural Eng 2018 Dec 17;15(6):066007. [doi: 10.1088/1741-2552/aad1a8] [Medline: 30221624]
23. Liu H, Yang Z, Meng F, Guan Y, Ma Y, Liang S, et al. Preoperative Heart Rate Variability as Predictors of Vagus Nerve Stimulation Outcome in Patients with Drug-resistant Epilepsy. Sci Rep 2018 Mar 01;8(1):3856. [doi: 10.1038/s41598-018-21669-3] [Medline: 29497072]
24. Schoukens J, Ljung L. Nonlinear System Identification: A User-Oriented Road Map. IEEE Control Systems Magazine 2019;39(6):28-99.

25. van Overschee P, de Moor B. Subspace Identification for Linear Systems. Norwell, MA: Kluwer Academic Publishers; 1996.
26. Ljung L. On The Consistency of Prediction Error Identification Methods. Math Sci Eng 1976;126:121-164. [doi: 10.1016/S0076-5392(08)60871-1]
27. Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. New York, NY: Springer-Verlag; 2002.
28. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974 Dec;19(6):716-723. [doi: 10.1109/tac.1974.1100705]
29. Gurel NZ, Gazi AH, Scott KL, Wittbrodt MT, Shah AJ, Vaccarino V, et al. Timing Considerations for Noninvasive Vagal Nerve Stimulation in Clinical Studies. 2019 Presented at: AMIA Annual Symposium; 2019; Washington, DC p. 1061-1070.
30. Mukkamala R, Cohen RJ. A forward model-based validation of cardiovascular system identification. Am J Physiol Heart Circ Physiol 2001 Dec;281(6):H2714-H2730 [FREE Full text] [doi: 10.1152/ajpheart.2001.281.6.H2714] [Medline: 11709441]
31. Ljung L. System Identification Toolbox User's Guide. URL: https://www.mathworks.com/help/pdf_doc/ident/ident_ug.pdf [accessed 2020-05-20]
32. Tikhonov AN. Solution of Incorrectly Formulated Problems and the Regularization Method. Soviet Math Doklady 1963;4:1035-1038. [doi: 10.1007/978-1-4612-5280-1_1]
33. Yang Y, Sani OG, Chang EF, Shanechi MM. Dynamic network modeling and dimensionality reduction for human ECoG activity. J Neural Eng 2019 Aug 14;16(5):056014. [doi: 10.1088/1741-2552/ab2214] [Medline: 31096206]
34. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis, Forecasting, and Control. Hoboken, NJ: John Wiley & Sons; 2016.
35. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. Melbourne, Australia: OTexts; 2018. URL: https://otexts.com/fpp2/ [accessed 2020-05-20]
36. Ljung L. System Identification: Theory for the User. In: Kailath T, editor. Prentice Hall Information and System Sciences Series. Upper Saddle River, NJ: Prentice-Hall, Inc; 1999.
37. Burger AM, D'Agostini M, Verkuil B, Van Diest I. Moving beyond belief: A narrative review of potential biomarkers for transcutaneous vagus nerve stimulation. Psychophysiology 2020 Jun 23;57(6):e13571. [doi: 10.1111/psyp.13571] [Medline: 32202671]
38. Clancy JA, Mary DA, Witte KK, Greenwood JP, Deuchars SA, Deuchars J. Non-invasive vagus nerve stimulation in healthy humans reduces sympathetic nerve activity. Brain Stimul 2014 Nov;7(6):871-877. [doi: 10.1016/j.brs.2014.07.031] [Medline: 25164906]
39. Khoo MCK. Physiological Control Systems: Analysis, Simulation, and Estimation. Piscataway, NJ: IEEE Press; 2018.
40. Ganzer P, Sharma G. Opportunities and challenges for developing closed-loop bioelectronic medicines. Neural Regen Res 2019 Jan;14(1):46-50 [FREE Full text] [doi: 10.4103/1673-5374.243697] [Medline: 30531069]
41. Smith OJM. A controller to overcome dead time. ISA J 1959;6:28-33.
42. Badran BW, Mithoefer OJ, Summer CE, LaBate NT, Glusman CE, Badran AW, et al. Short trains of transcutaneous auricular vagus nerve stimulation (taVNS) have parameter-specific effects on heart rate. Brain Stimul 2018 Jul;11(4):699-708 [FREE Full text] [doi: 10.1016/j.brs.2018.04.004] [Medline: 29716843]
43. George MS, Nahas Z, Bohning DE, Mu Q, Andrew Kozel F, Borckhardt J, et al. Mechanisms of action of vagus nerve stimulation (VNS). Clin Neurosci Res 2004 Jul;4(1-2):71-79. [doi: 10.1016/j.cnr.2004.06.006]
44. Qing KY, Wasilczuk KM, Ward MP, Phillips EH, Vlachos PP, Goergen CJ, et al. B fibers are the best predictors of cardiac activity during Vagus nerve stimulation: Qing, vagal B fiber activation and cardiac effects. Bioelectron Med 2018 Apr 20;4(1):5 [FREE Full text] [doi: 10.1186/s42234-018-0005-8] [Medline: 32232081]
45. Harel T, Gath I, Ben-Haim SA. System response of the sinoatrial node during vagal stimulation. Physiol Meas 1998 May 01;19(2):149-157. [doi: 10.1088/0967-3334/19/2/002] [Medline: 9626679]
46. Nonis R, D'Ostilio K, Schoenen J, Magis D. Evidence of activation of vagal afferents by non-invasive vagus nerve stimulation: An electrophysiological study in healthy volunteers. Cephalalgia 2017 Nov 26;37(13):1285-1293 [FREE Full text] [doi: 10.1177/0333102417717470] [Medline: 28648089]
47. Mourdoukoutas AP, Truong DQ, Adair DK, Simon BJ, Bikson M. High-Resolution Multi-Scale Computational Model for Non-Invasive Cervical Vagus Nerve Stimulation. Neuromodulation 2018 Apr 27;21(3):261-268 [FREE Full text] [doi: 10.1111/ner.12706] [Medline: 29076212]
48. Tyler WJ, Wyckoff S, Hearn T, Hool N. The Safety and Efficacy of Transdermal Auricular Vagal Nerve Stimulation Earbud Electrodes for Modulating Autonomic Arousal, Attention, Sensory Gating, and Cortical Brain Plasticity in Humans. bioRxiv 2019 [FREE Full text] [doi: 10.1101/732529]
49. McDowall D, McCleary R, Bartos BJ. Interrupted Time Series Analysis. New York, NY: Oxford University Press; 2019.
50. Chon K, Mukkamala R, Toska K, Mullen T, Armoundas A, Cohen R. Linear and nonlinear system identification of autonomic heart-rate modulation. IEEE Eng Med Biol Mag 1997;16(5):96-105. [doi: 10.1109/51.620500] [Medline: 9313086]

Abbreviations

AICc: small sample size-corrected Akaike information criterion
AU: arbitrary units
ECG: electrocardiogram
FDA: Food and Drug Administration
PPG: photoplethysmogram
RMSE: root mean square error
tcVNS: transcutaneous cervical vagus nerve stimulation

Edited by G Eysenbach; submitted 20.05.20; peer-reviewed by IV George, O Joel; comments to author 26.06.20; revised version received 27.06.20; accepted 26.07.20; published 22.09.20.

Please cite as:
Gazi AH, Gurel NZ, Richardson KLS, Wittbrodt MT, Shah AJ, Vaccarino V, Bremner JD, Inan OT
Digital Cardiovascular Biomarker Responses to Transcutaneous Cervical Vagus Nerve Stimulation: State-Space Modeling, Prediction, and Simulation
JMIR Mhealth Uhealth 2020;8(9):e20488
URL: http://mhealth.jmir.org/2020/9/e20488/
doi:10.2196/20488
PMID:32960179

Original Paper

Using Machine Learning and Smartphone and Smartwatch Data to Detect Emotional States and Transitions: Exploratory Study

Madeena Sultana[1,2*], PhD; Majed Al-Jefri[1,3*], PhD; Joon Lee[1,2,4], PhD

[1]Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[4]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[*]these authors contributed equally

Corresponding Author:
Joon Lee, PhD
Data Intelligence for Health Lab
Cumming School of Medicine
University of Calgary
Teaching Research & Wellness 5E17
3280 Hospital Dr. NW
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 220 2968
Email: joonwu.lee@ucalgary.ca

Abstract

Background: Emotional state in everyday life is an essential indicator of health and well-being. However, daily assessment of emotional states largely depends on active self-reports, which are often inconvenient and prone to incomplete information. Automated detection of emotional states and transitions on a daily basis could be an effective solution to this problem. However, the relationship between emotional transitions and everyday context remains to be unexplored.

Objective: This study aims to explore the relationship between contextual information and emotional transitions and states to evaluate the feasibility of detecting emotional transitions and states from daily contextual information using machine learning (ML) techniques.

Methods: This study was conducted on the data of 18 individuals from a publicly available data set called ExtraSensory. Contextual and sensor data were collected using smartphone and smartwatch sensors in a free-living condition, where the number of days for each person varied from 3 to 9. Sensors included an accelerometer, a gyroscope, a compass, location services, a microphone, a phone state indicator, light, temperature, and a barometer. The users self-reported approximately 49 discrete emotions at different intervals via a smartphone app throughout the data collection period. We mapped the 49 reported discrete emotions to the 3 dimensions of the pleasure, arousal, and dominance model and considered 6 emotional states: discordant, pleased, dissuaded, aroused, submissive, and dominant. We built general and personalized models for detecting emotional transitions and states every 5 min. The transition detection problem is a binary classification problem that detects whether a person's emotional state has changed over time, whereas state detection is a multiclass classification problem. In both cases, a wide range of supervised ML algorithms were leveraged, in addition to data preprocessing, feature selection, and data imbalance handling techniques. Finally, an assessment was conducted to shed light on the association between everyday context and emotional states.

Results: This study obtained promising results for emotional state and transition detection. The best area under the receiver operating characteristic (AUROC) curve for emotional state detection reached 60.55% in the general models and an average of 96.33% across personalized models. Despite the highly imbalanced data, the best AUROC curve for emotional transition detection reached 90.5% in the general models and an average of 88.73% across personalized models. In general, feature analyses show that spatiotemporal context, phone state, and motion-related information are the most informative factors for emotional state and transition detection. Our assessment showed that lifestyle has an impact on the predictability of emotion.

Conclusions: Our results demonstrate a strong association of daily context with emotional states and transitions as well as the feasibility of detecting emotional states and transitions using data from smartphone and smartwatch sensors.

(JMIR Mhealth Uhealth 2020;8(9):e17818) doi:10.2196/17818

KEYWORDS

mHealth; mental health; emotion detection; emotional transition detection; spatiotemporal context; supervised machine learning; artificial intelligence; mobile phone; digital biomarkers; digital phenotyping

Introduction

The emotional states of individuals may change frequently over time. Research has demonstrated the potential of recording daily emotional states and moods in health and well-being, including the early diagnosis of mental illness and disorders [1-3]. However, the process of recording emotional states and moods largely depends on active self-reports less frequently than daily. However, with the unprecedented rise of smartphones and wearable devices as well as the advancement in built-in sensors within these devices, it is possible to passively collect multimodal data from people's everyday lives at a much higher frequency. The self-reporting problem of personal health tracking can therefore be solved to a great extent by leveraging machine learning (ML) algorithms on the myriad of data collected by smartphones and wearables.

Predicting and monitoring mental health illnesses and diseases such as depression, bipolar disorder, Alzheimer disease, and schizophrenia via smartphones and wearable sensors have been an active area of research over the last few years. Research has been conducted in the quest for gold standard digital biomarkers that can be collected through consumer-grade smartphones and wearable sensors (eg, accelerometer, audio, location, phone log, sound features, etc) to detect mental health disorders in the early stages [4,5]. It is evident that mobility patterns, location variations, and phone usage patterns captured by smartphones can aid in identifying patients with mental health illnesses and disorders [1,2,6-9]. Early detection of depressive symptoms by applying deep neural networks and ML techniques to self-reported contextual data through smartphones obtained promising results [8]. However, the aforementioned apps were designed from disease and illness perspectives and did not consider the automated detection of regular emotional states and transitions in everyday life.

The association between everyday mood, emotion, and well-being and sensed data via smartphones and wearables has been studied recently. For example, Helbich [10] found an association of people's mental well-being with the neighborhood they live in, the places they visit, and the environmental exposure they experience. In a similar study by Sandstrom et al [11], subjects reported emotional pleasantness in a societal environment, whereas positive and negative arousal at work. A daily mood assessment tool was proposed by Ma et al [12] that utilized mobile phone sensor data such as location, audio, text messages, accelerometer, and light to classify mood. However, this study considered limited contextual parameters, and the subjective variability as well as transitions of emotional states remained uninvestigated. The prediction of the Ecological Momentary Assessment scores from smartphone data such as text messages, screen time, app usage, accelerometer, and phone camera have been studied by Asselbergs et al [13], who reported a promising but lower prediction accuracy than naive benchmark approaches. Studies have also shown an association between mobile sensor data (eg, phone usage, motion, conversation, mobility, screen time, and skin conductance) and the academic performance and mental health conditions of college students [14,15]. Budner et al [16] classified 9 emotional states in 2 dimensions (pleasure and activation) of the circumplex model by applying a random forest on the smartwatch sensor data such as motion, heart rate, light level, GPS coordinates, day of the week, humidity, air pressure, cloudiness, and windiness. In a similar study, an ML-based model was proposed by Zhang et al [17] to recognize compound emotional states in pleasure and arousal dimensions from smartphone data (eg, microphone, accelerometer, GPS, text message, phone call, app usage). Promising results have been obtained in some recent works for daily mood and stress forecasting [3,18,19], where deep neural networks were applied to physiological, personality traits, and sensed data acquired from a large-scale global population using smartphones and wearables.

Despite encouraging results and progress, gaps in the literature include the lack of research on the association between emotional transition, sensed data, and contextual information; subjective variability in classification performance; and feasibility of frequent emotional state and transition detection. In addition, the majority of the previous studies are based on the circumplex model of affect [20], which considers only the pleasure and arousal dimensions (also known as *core affect*). However, research has demonstrated that considering all the 3 dimensions can facilitate a better understanding and interpretation of persons' emotional states [21]. Therefore, the third dimension of emotional states—dominance—needs to be included in emotion recognition research.

In this study, we aim to fill the aforementioned gaps. Our main objective is to study the feasibility of detecting emotional states and transitions every 5 min by applying ML to the data acquired from smartphone and smartwatch sensors. Our study includes all the 3 dimensions of emotional states (pleasure, arousal, and dominance [PAD]) as well as the variability of interpersonal data. The remainder of this paper is organized as follows. The Methods section describes the methodology followed in our study along with an overview of the data set description and preparation. The Results section shows the results obtained for emotion transition and detection tasks for both general and personalized models. The Discussion section presents the results with a deeper analysis of the features.

Methods

The Data Set

We obtained data from a publicly available data set called ExtraSensory [22]. This data set was collected by the researchers of the University of California, San Diego (UCSD), in 2015 to 2016 for automated context labeling from signals captured via

a wide range of smartphone and smartwatch sensors such as an accelerometer, a gyroscope, a magnetometer, a compass, location services, audio, phone state, light, air pressure, humidity, and temperature [23]. It contains data from 60 subjects in free-living conditions, who were mainly students (both undergraduate and graduate) and research assistants at UCSD. The sensor data were collected every minute, and the contextual data were self-reported at different intervals by the users. This data set also contains optionally self-reported discrete emotions at different time intervals. There were a total of 49 different discrete emotions (eg, active, calm, happy, sleepy, etc) that were reported by the subjects and the interval varied from 1 min to several days. Researchers processed and cleaned the self-reported data by combining various sources of information such as location and other labels [23] to make them reliable. Both the raw and cleaned versions of self-reported data are available. We used the cleaned version in this study.

The Pleasure, Arousal, and Dominance Model

The PAD model was developed by Mehrabian and Russell [24] in 1974 to assess individuals' psychological responses to environmental perception and experience. Persons' emotional states can be perceived in 3 basic dimensions: pleasure, arousal, and dominance. Pleasure is the dimension of positive or negative feelings [24]. Arousal represents states of mental responsiveness [25]. Dominance is the perceptual cognitive dimension of the feeling influenced or controlled [25]. Our study includes all the 3 dimensions of emotional states.

Data Preparation

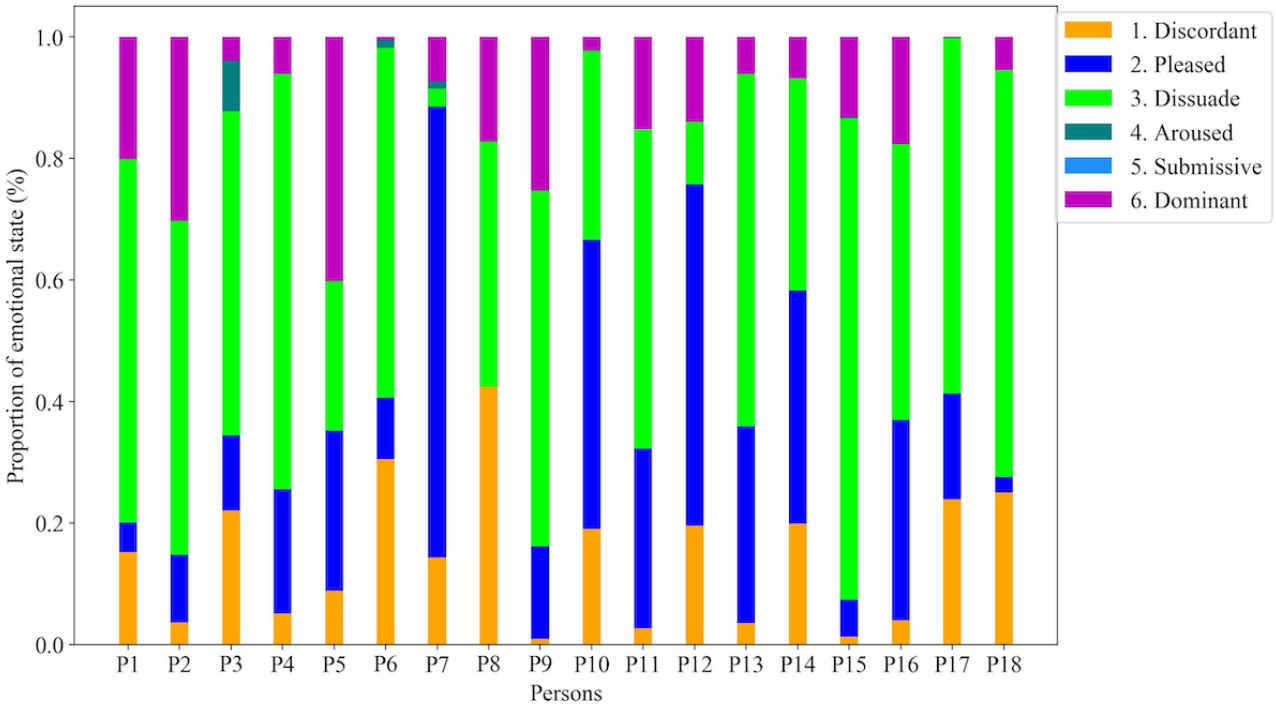
Inclusion and Exclusion

In this exploratory study, we aimed to apply ML in the 2 setups. First, we built personalized models using each person's data to analyze the impact of variability across individuals. Second, we built generalized models using data from multiple individuals and validated them using data from other individuals who were left out during training. Although 37 subjects in the ExtraSensory [22] data set reported their emotions at some points during data collection, only 18 of them had more than 1000 samples and less than 90% missing data. Therefore, we considered these 18 subjects in our study. We considered all signals collected from the smartphone and smartwatch sensors, timestamps, latitudes, and longitudes as features.

Affective Ratings of Emotions

In this study, we used the Affective Norms for English Words (ANEW) [26] to map the 49 discrete emotions to the PAD model. ANEW was developed by the Center for the Study of Emotion and Attention to provide standardized materials to researchers studying emotion and attention. The latest ANEW database [27] contains affective meanings of nearly 14,000 English lemmas rated by a larger cohort of 1827 participants with a wide range of diversities, including age, occupation, and educational differences. We used the latest database to map the 49 emotions to the 3 ratings of pleasure (*p*), arousal (*a*), and dominance (*d*). Therefore, each linguistic emotion label was converted into 3 continuous values on a scale of 1 to 9, where 1 and 9 indicate the lowest and highest intensity, respectively, in the corresponding PAD dimension. A list of the 49 emotions used in this study with their corresponding PAD values is included in Multimedia Appendix 1 [22,23,28]. We scaled the emotional ratings (R_s) in the range of –4 to +4 by subtracting 5. Then, we considered 6 states of emotions on the basis of the intensity (sign) in the 3 dimensions: discordant, pleased, dissuaded, aroused, submissive, and dominant. We calculated the prevailing emotional state at any point of time for a person by considering the absolute maximum value of (R_s) and its sign. The absolute maximum value indicates the dimension, whereas the sign represents the direction. Therefore, the emotional state at any point represents which of the 3 dimensions is prevailing and in what direction. For example, the emotional rate (R_s) of *happy* is 8.47, 6.05, and 7.21 *for p*, *a*, and *d*, respectively. The corresponding scaled values will be 3.47, 1.05, and 2.21, respectively. Here, the 3 values are positive, and the prevailing emotional state is (*+p)* pleased. Therefore, this emotional state will be assigned to the class *pleased.* Similarly, *angry* (*p*=2.53, *a*=6.2, and *d*=4.11) will be scaled to –2.47, 1.2, and –0.89, respectively, with a maximum absolute value of 2.47 and in the negative direction. Hence, *angry* will be assigned to the class *discordant*. There was 1 case (the emotion *interested*), *where p* and *a* were equal. In this case, we had 2 dominant dimensions, and we chose the first positive value (in the order of *PAD*), *p*, to represent the dominant emotion. Although this is a limitation, there was only 1 emotion *interested* that had 2 equal values, and there were few cases with this emotion in the data set compared with the other 48 emotions. Although we considered 6 categories of emotional states, not all classes were present in every person's data. Depending on the person, 1 or 2 emotional states were absent in the data set. Figure 1 shows the different emotional states present in each person's data.

Figure 1. Proportion (%) of 6 emotional states per person. Pleased and dissuaded are the most frequent, whereas submissive and aroused are the least frequent emotional states among the 18 persons in our data set.



Feature Engineering

Initially, we merged all features from sensors, location data, and self-reported contextual information. Sensor measurements were recorded for 20 seconds every minute, and the data collection period varied from 3 to 9 days for each person. The number of samples for each person varied from 1164 to 6263. The data set contains a mixture of binary and continuous variables. We also engineered 7 additional temporal and spatial features from timestamps and location data. Overall, the features can be categorized as follows.

Motion

We considered 138 features calculated from the raw measurements from 3 smartphone sensors (an accelerometer, a gyroscope, and a magnetometer) and 2 smartwatch sensors (an accelerometer and a compass). These are continuous variables.

Audio

We considered 28 naive features calculated as the averages and standard deviations of the 13 Mel Frequency Cepstral Coefficients from the approximately 20-second recording window and the overall power of the audio.

Location

We considered 17 location features measured from the relative locations and movement variability in every minute of persons. We also feature engineered 3 more location features: *cl_latitude*, *cl_longitude*, and *geo_dist*. We clustered the neighboring latitudes and longitudes using geohash [29]. Geohash is a geocoding system invented by Niemeyer [29] that enables the grouping of neighboring points in a rectangular cell defined by a precision value. We used a precision value of 8 to cluster the neighboring latitude and longitude within 38.2 m × 19.1 m. The rectangular box worked as a bounding box for all neighboring spatial points falling into this area. After geohashing, the geocodes were decoded back to clustered latitude (*cl_latitude*) and longitude (*cl_longitude*) values. We calculated the *geo_dist* feature as the Haversian distance traveled by the person since the previous time stamp.

Phone State

We considered 28 binary features that indicate the sensed state of the phone, such as app states, battery plugged, battery states, ringer mode, on the phone, Wi-Fi status, screen brightness, and battery level.

Environmental

The ExtraSensory data set also contains 6 environmental variables such as light, pressure, humidity, and temperature. All of these continuous variables were included in our primary feature list. However, there were many missing values for these features because not all phones had all the sensors.

Temporal

We engineered 5 variables from the recorded time stamps to explore the temporal pattern of emotional states and transitions: *minute of the hour*, *minute of the day*, *hour of the day*, *day of week*, and *time difference in minutes*. As the data set was very sparse, we calculated the *time difference in minutes* variable to measure how many minutes elapsed since the last record. The remaining 4 variables were categorical variables.

Contextual

We also considered 51 binary contextual labels such as indoor, outdoor, eating, and in a car, which were self-reported by the subjects at various intervals. We assumed that this self-reported information was correct in all cases to focus on automatically recognizing the dominant emotion without dealing with noisy estimates of the context. Although the latest ExtraSensory app

[28] is capable of recognizing contextual information passively on the basis of raw sensor data, the data set used in this study did not include the output from this new feature.

A complete list of features is included in Multimedia Appendix 1.

Data Resampling, Cleaning, and Imputation

Our study aims to detect emotional transitions and states in small time intervals. Therefore, we resampled all data to a frequency of every 5 min. In the original data, the number of samples in 5-min intervals varied from 0 to 5. During resampling, we calculated the average of all continuous variables, the summation of all binary variables, and the maximum of all ordinal variables for all samples within the 5-min interval. This allowed us to have an evenly spaced sampling frequency over time and reduced missing data. All missing values were replaced by a large negative number to indicate missingness. Features were standardized by removing the mean and scaling to unit variance. This was done on the basis of the training sets.

Ethics Approval

As ExtraSensory is a public data set, research ethics approval was waived.

Emotional Transition and State Detection

Feature Handling

For emotional transition detection, we considered the changes in features from the previous window. Therefore, the feature set $T_{t,k}$ at any time t was calculated as follows:

In equation 1, the total number of features is n, $f_{t,k}$ represents the value of the kth feature at the tth window, and $f_{t-1,k}$ represents the value of the kth feature at the $(t-1)$th window. The intuition was to feed the ML models with the changes in information to find any pattern associated with changes in the captured data. The original form of the resampled features was used for emotional state detection.

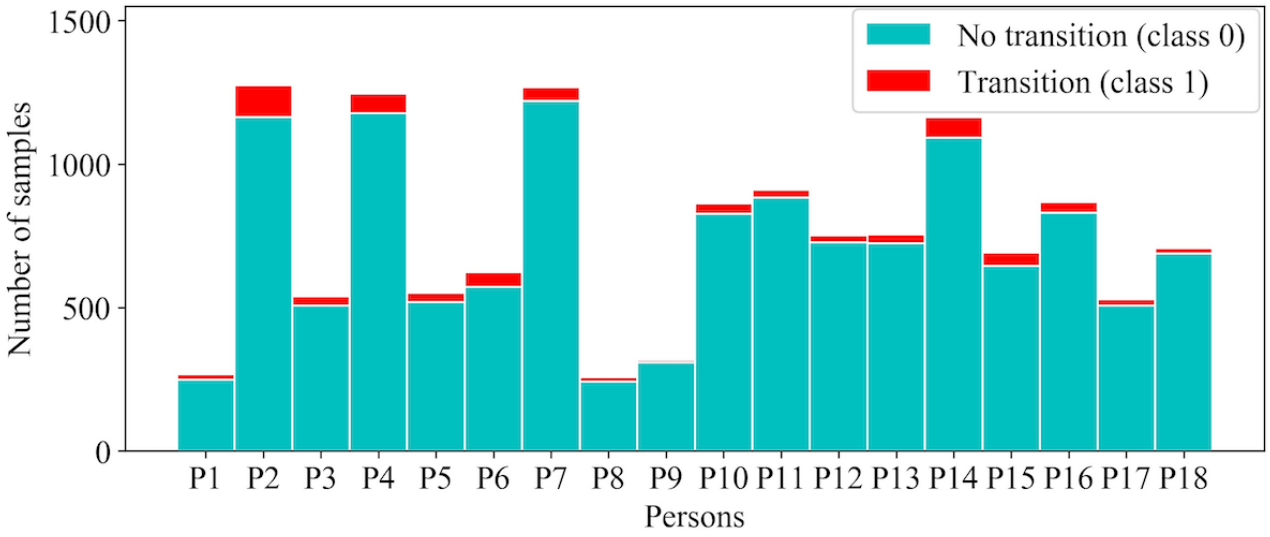
Next, we analyzed the features to select a smaller set of important features. We selected the k best features for each model by applying the SelectKBest feature selection function from the feature_selection package provided by sklearn, where the following values were experimented for k: 50, 70, 90, and 110. This feature selection process was applied independently for emotional transition and state detection and for general and personalized models. This resulted in a different number of features for each model. Additionally, columns with more than 30% missing data were removed. Location data were removed from the general models to make the models as generalizable as possible.

Machine Learning Models

We developed general models for all individuals as well as personalized models for each person to explore the impact of interpersonal variability on the performance of emotional transition and state detection.

We used 5 supervised ML algorithms: logistic regression (LR), random forest (RF), XGBoost (XGB), CatBoost (CB), and multilayer perceptron. Emotional transition detection is a binary classification problem, where 0 and 1 denote no change and change in emotional state, respectively, over the last 5-min window. Owing to the sparsity of the data, the target variable was overly imbalanced. Figure 2 shows the class imbalance of emotional state transitions of the 18 persons. Hence, we also applied 2 imbalance handling techniques that we explain in the *Imbalance Handling* section below.

Figure 2. A stacked bar chart of the distribution of the 2 classes (transition and no transition) per person.



Emotional state detection is a multiclass classification problem, where we intended to classify the prevailing emotional state of a person at a given time into one of the following 6 classes: 1, discordant; 2, pleased; 3, dissuaded; 4, aroused; 5, submissive; and 6, dominant. However, as shown in Figure 1, the emotional state classes were also imbalanced. We removed any class from a person's data having less than or equal to 6 samples. As a result, the number of classes varied from 3 to 5 for each person.

Hyperparameters for each model were determined using a stratified cross-validated grid search over a parameter grid. In the general models, we used a six-fold, leave-3-people-out cross-validation, where for each fold, the models were trained on 15 individuals' data and tested on the remaining 3 data points. Hyperparameters were tuned by optimizing the F_1 score. The tuned hyperparameters are listed in Multimedia Appendix 1. For the personalized models, a five-fold, stratified cross-validation was used to fine-tune the hyperparameters and select the best-performing models. The total number of samples per person varied from 257 to 1268. We measured 7 performance metrics to evaluate the classification performance: accuracy, balanced accuracy, precision, recall, F_1 score, specificity, and area under the receiver operating characteristic (AUROC) curve. For emotional state detection, the macro precision, recall, F_1 score, specificity, and AUROC curve were measured to emphasize the detection performance for the minority classes.

Imbalance Handling

In this study, Synthetic Minority Over-sampling Technique (SMOTE) and Support Vector Machines Synthetic Minority Over-sampling Technique (SVMSMOTE) [30] were applied to mitigate class imbalance. These are oversampling methods that create synthetic data of the minority classes to decrease the imbalance. All imbalance handling techniques were applied only on training data to avoid data leakage between training and test sets.

Feature Analysis

We analyzed the importance of the 7 categories of features used in this study for both emotional transition and state detection. The detection performance of emotional transition and state varied for different persons, which we categorized as best, average, and worst performances. We used the output of the XGB classifier for the feature importance analysis to explore the best-performing features.

Software

This study was conducted in Python 3 with the following packages: Scikit-Learn (0.22), CB, XGB, and SHapley Additive exPlanation (SHAP). Python codes are publicly available on GitHub [31].

Results

A summary of the data set containing the number of days and the percentage of missing data is presented in Table 1. Table 1 shows that the average amount of missing values in the data set was approximately 63%, where the range varied from 38.66% to 88.8% for different persons.

Table 1. Summary of the 18 persons' (P1-P18) data used in this study.

Person	Number of days	Data (5-min window)	
		Windows with complete data, n	Missing data[a] (%)
P1	3	268	68.9
P2	8	1275	44.62
P3	7	539	73.21
P4	8	1245	45.92
P5	8	551	76.04
P6	8	623	72.92
P7	8	1268	44.92
P8	8	257	88.80
P9	7	317	84.23
P10	8	862	62.54
P11	8	911	60.42
P12	8	752	67.32
P13	9	755	70.83
P14	7	1164	42.21
P15	8	692	69.92
P16	7	868	56.89
P17	3	529	38.7
P18	7	707	64.88
Average (SD)	7.22 (1.63)	754.61 (328.06)	62.96 (14.71)

[a]Missing value is the percentage of missing windows (time slots).

Emotional Transition or State Detection Results

In terms of general models for emotional transition detection, the best-performing classifiers were LR, XGB, and CB. Table 2 shows the results of these classifiers. As shown in Table 2, the best average AUROC curve of 90.5% was obtained by LR, followed by XGB and CB with 89.72% and 89.24%, respectively. The table also shows that using imbalance handling improved average recall only but did not improve the results in general. In general, SVMSMOTE produced better results in terms of imbalance handling than SMOTE for most models. Complete results including the results using SMOTE are shown in the Multimedia Appendix 1.

Table 2. Results of the general models for emotional transition detection on the basis of a six-fold, leave-3-people-out cross-validation.

Characteristics	Machine learning models[a,b] (without imbalance handling)			Imbalance handling using Support Vector Machines Synthetic Minority Over-sampling Technique		
	LR[c]	XGB[d]	CB[e]	LR	XGB	CB
Accuracy, % (SD)	*94.77 (1.64)*	94.50 (1.54)	94.60 (1.58)	91.16 (1.24)	5.33 (1.46)	41.31 (33.75)
Balanced accuracy, % (SD)	60.59 (4.26)	*66.92 (5.41)*	65.02 (5.47)	80.76 (4.68)	50.09 (0.18)	60.56 (10.45)
Precision (macro), % (SD)	50.08 (10.55)	*51.77 (16.98)*	51.01 (12.13)	32.59 (5.47)	5.17 (1.58)	9.37 (4.73)
Recall (macro), % (SD)	22.32 (8.49)	36.03 (11.70)	31.98 (12.32)	69.05 (9.46)	*100.00 (0.00)*	82.27 (20.45)
Specificity (macro), % (SD)	*98.85 (0.44)*	97.81 (1.29)	98.05 (1.53)	92.48 (0.99)	0.17 (0.36)	38.84 (36.35)
AUROC[f] (macro), % (SD)	*90.50 (3.01)*	89.72 (2.51)	89.24 (2.51)	90.26 (3.20)	60.49 (14.11)	74.77 (9.75)
F_1 (macro), % (SD)	29.89 (9.51)	38.85 (7.83)	36.46 (6.72)	*43.63 (5.05)*	9.80 (2.83)	15.78 (6.53)

[a]Average (SD) across six-fold.

[b]The highest value of each metric is italicized.

[c]LR: logistic regression.

[d]XGB: XGBoost.

[e]CB: CatBoost.

[f]AUROC: area under the receiver operating characteristic.

Table 3 reports the performance measures for the general models for emotional state detection. As shown in Table 3, the best results were obtained from LR, CB, and RF. In particular, LR achieved the best average AUROC curve of 60.23%. Adding imbalance handling slightly improved some of the metrics such as specificity and balanced accuracy. The full results from all models and imbalance handling techniques can be found in Multimedia Appendix 1.

Table 3. Results of the general models for emotional state detection on the basis of a six-fold, leave-3-people-out cross-validation.

Characteristics	Machine learning models[a,b] (without imbalance handling)			Imbalance handling using Support Vector Machines Synthetic Minority Over-sampling Technique		
	LR[c]	CB[d]	RF[e]	LR	CB	RF
Accuracy, % (SD)	40.60 (9.50)	*44.10 (13.93)*	44.04 (14.42)	32.61 (2.65)	39.52 (13.24)	38.99 (8.71)
Balanced accuracy, % (SD)	22.83 (2.64)	24.32 (3.23)	21.84 (2.03)	*30.66 (8.82)*	23.32 (2.91)	23.27 (2.43)
Precision (macro), % (SD)	*34.25 (5.26)*	29.52 (11.39)	20.36 (7.50)	26.48 (2.32)	27.18 (10.23)	25.33 (2.20)
Recall (macro), % (SD)	*38.86 (6.68)*	33.29 (8.57)	29.40 (3.87)	28.27 (7.76)	27.81 (5.77)	25.94 (2.08)
Specificity (macro), % (SD)	63.37 (12.40)	72.40 (3.62)	71.23 (4.73)	*82.93 (0.76)*	75.36 (4.66)	76.93 (2.63)
AUROC[f] (macro), % (SD)	60.23 (8.15)	58.58 (6.97)	55.21 (4.31)	*60.55 (3.41)*	56.83 (6.14)	55.43 (4.37)
F_1 (macro), % (SD)	*30.60 (5.83)*	25.45 (10.72)	19.20 (5.19)	23.04 (2.82)	21.20 (6.95)	21.81 (3.24)

[a]Average (SD) across six-fold (average value for each metric).

[b]The highest value of each metric is italicized.

[c]LR: logistic regression.

[d]CB: CatBoost.

[e]RF: random forest.

[f]AUROC: area under the receiver operating characteristic.

For the personalized emotional transition detection models, all models performed quite well in general, and it was not possible to pinpoint 1 single best ML model and imbalance handling technique for all the 18 persons. Table 4 reports the performance

measures obtained from RF, XGB, and CB. The standard deviations of the measures indicate large variabilities across 18 persons. The best measures highlighted in Table 4 demonstrate the variabilities of the measures across different ML models. As shown in Table 4, the best average AUROC curve of 88.01% was obtained by RF without imbalance handling, whereas SVMSMOTE helped improve recall in general (especially in XGB) and produced the best average AUROC curve of 88.7% with CB. The detailed results of all classifiers and imbalance handling techniques are listed in the Multimedia Appendix 1.

Table 4. Results of the personalized models for emotional transition detection on the basis of the 5-fold, stratified cross-validation.

Characteristics	Machine learning models[a,b] (without imbalance handling)			Imbalance handling using Support Vector Machines Synthetic Minority Over-sampling Technique		
	RF[c]	XGB[d]	CB[e]	RF	XGB	CB
Accuracy, % (SD)	93.49 (4.73)	*94.82 (2.55)*	94.29 (3.17)	90.31 (7.17)	89.54 (6.06)	92.34 (5.01)
Balanced accuracy, % (SD)	66.88 (10.08)	65.91 (7.93)	66.89 (8.97)	66.80 (5.93)	*75.86 (8.52)*	70.28 (8.06)
Precision (macro), % (SD)	49.97 (21.23)	*52.34 (17.22)*	48.67 (18.09)	34.29 (17.13)	35.69 (13.32)	42.37 (14.18)
Recall (macro), % (SD)	37.12 (20.78)	33.71 (15.69)	36.45 (18.10)	40.77 (13.49)	*60.76 (16.37)*	45.82 (15.17)
Specificity (macro), % (SD)	96.65 (4.91)	*98.12 (1.42)*	97.34 (2.33)	92.82 (7.63)	90.97 (6.03)	94.75 (4.53)
AUROC[f] (macro), % (SD)	88.01 (5.67)	87.84 (6.81)	87.62 (7.04)	85.08 (7.20)	87.74 (6.63)	*88.73 (6.24)*
F_1 (macro), % (SD)	36.40 (17.87)	38.00 (15.92)	38.34 (17.07)	32.92 (12.48)	*41.85 (13.46)*	40.44 (13.88)

[a]Average (SD) across 18 persons.

[b]The highest value of each metric is italicized.

[c]RF: random forest.

[d]XGB: XGBoost.

[e]CB: CatBoost.

[f]AUROC: area under the receiver operating characteristic.

Compared with the performance of the general models for the emotional state detection task, the performance of the personalized models was substantially better. Table 5 reports the performance measures obtained from CB, XGB, and RF. As shown in Table 5, the best average AUROC curve of 96.33% was obtained by CB followed by XGB and then RF. Applying imbalance handling techniques slightly improved the balanced accuracy, recall, and specificity. While all classes were maintained in the general models, the number of classes varied between 4 and 5 across the personalized models. Complete results including those from SMOTE are shown in the Multimedia Appendix 1.

Table 5. Results of the personalized models for emotional state detection on the basis of a five-fold, stratified cross-validation.

Characteristics	Machine learning models[a,b] (without imbalance handling)			Imbalance handling using Support Vector Machines Synthetic Minority Over-sampling Technique		
	CB[c]	XGB[d]	RF[e]	CB	XGB	RF
Accuracy, % (SD)	*86.53 (8.08)*	82.73 (7.18)	80.92 (10.69)	85.27 (9.64)	81.54 (8.90)	78.21 (10.78)
Balanced accuracy, % (SD)	74.92 (13.55)	69.51 (12.18)	67.85 (14.10)	*77.73 (13.34)*	74.50 (12.03)	70.72 (13.92)
Precision (macro), % (SD)	*87.03 (9.37)*	82.49 (7.32)	82.11 (8.67)	84.44 (9.83)	77.71 (11.09)	75.94 (12.02)
Recall (macro), % (SD)	77.88 (11.35)	74.54 (7.70)	71.96 (10.87)	*80.00 (10.74)*	77.23 (10.36)	73.30 (11.91)
Specificity (macro), % (SD)	92.92 (4.80)	90.48 (5.64)	89.83 (6.58)	*93.15 (4.84)*	91.96 (4.71)	90.19 (5.67)
AUROC[f] (macro), % (SD)	*96.33 (3.26)*	94.81 (2.96)	93.74 (5.62)	96.26 (3.77)	94.51 (3.39)	92.93 (5.27)
F_1 (macro), % (SD)	*79.48 (11.38)*	74.87 (8.45)	72.46 (12.07)	79.47 (12.00)	74.51 (12.05)	70.73 (13.34)

[a]Average (SD) across 18 persons.

[b]The highest value of each metric is italicized.

[c]CB: CatBoost.

[d]XGB: XGBoost.

[e]RF: random forest.

[f]AUROC: area under the receiver operating characteristic.

Feature Analysis Results

Using the output of the XGB classifier for feature importance analysis, we explored the best-performing features. Figures 3 and 4 show the 20 most important features of the general model using XGB and SHAP [32] for emotional transition and state detection, respectively. The contextual information (prefix *label*) ranked higher among the 7 categories of features. Contextual information features appear among the top 4 features for both emotional transition and state detection. Other important features included motion, phone state, and temporal for both emotional transition and state detection.

Figure 3. Feature importance for emotional transition detection of the general model obtained using XGBoost and shapley additive explanation. The figure represents the contribution of the corresponding feature to detect whether there is a transition.

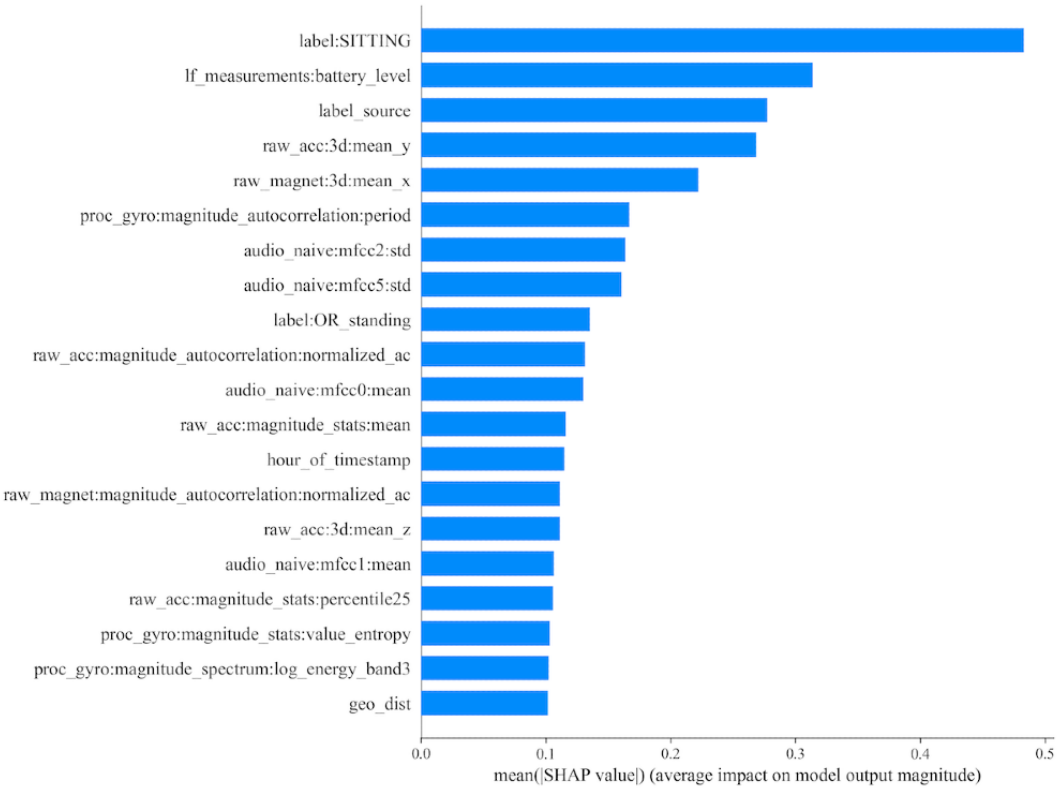
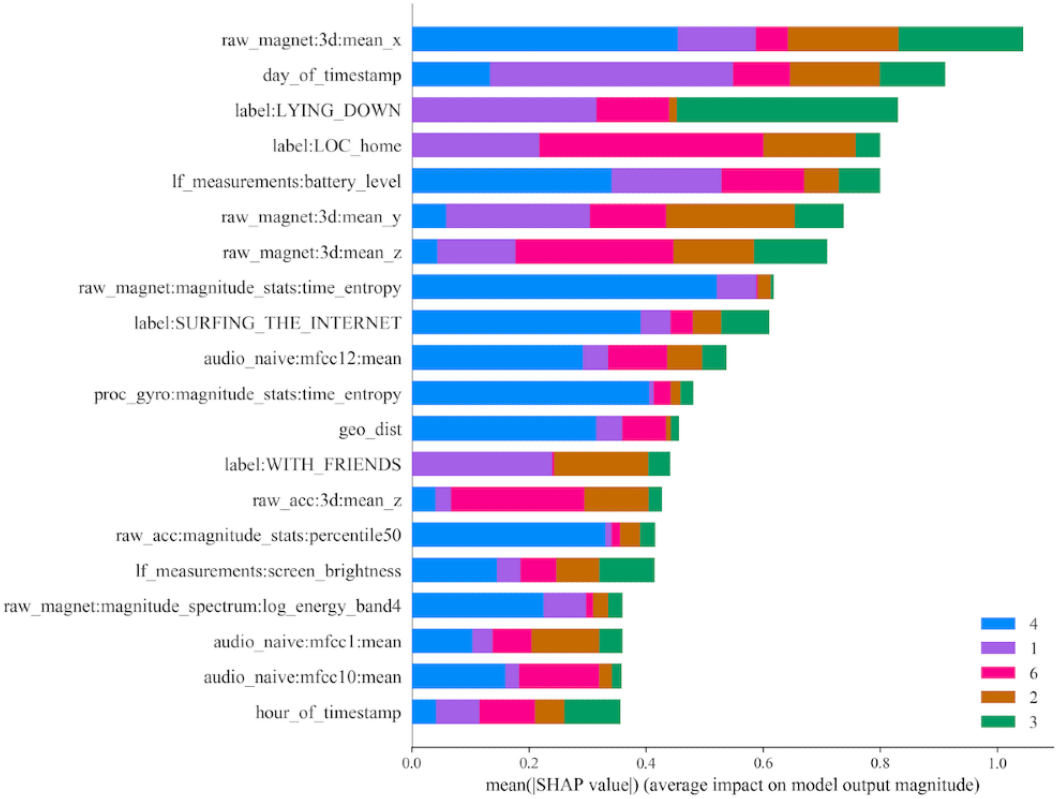


Figure 4. Feature importance for emotional state detection of the general model obtained using XGBoost and shapley additive explanation. The color-coded portions represent the contribution of the corresponding feature to detect different states (classes) of emotions: 1, discordant; 2, pleased; 3, dissuaded; 4, aroused; 5, submissive; and 6, dominant. SHAP: SHapley Additive exPlanation.



We also explored the influential feature categories by considering the 3 most important features across all 18 persons. Figures 5 and 6 demonstrate the importance of the 7 feature categories for emotional transition and state detection, respectively. This was done manually; each category is a bin of a set of features as described in the *Feature Engineering* section. Instead of the feature itself, we considered which bins the top-3 features belong to. This was done over folds for each person. For calculation, we considered a3 empty matrix, where 7 is for each category and 3 is for the top 3 ranks. Then we incremented the counter of the corresponding category and ranked the top 3 features of each fold for each person. For example, in Figure 3 (although for a general model but for the sake of explanation), the top 3 features are *label: SITTING*, *If_measurement:battery_level*, and *label: Source*, where the first and third features belong to the contextual category and the second feature falls under the phone state category. This will increase the first and third rank counters of the contextual category and the second rank counter of the phone state category. This was done based on the XGB classifier and SHAP.

Figure 5. Importance of feature categories for emotional transition detection. The color-coded portions of each category represent their contributions to the top 3 ranks of features of 18 persons.

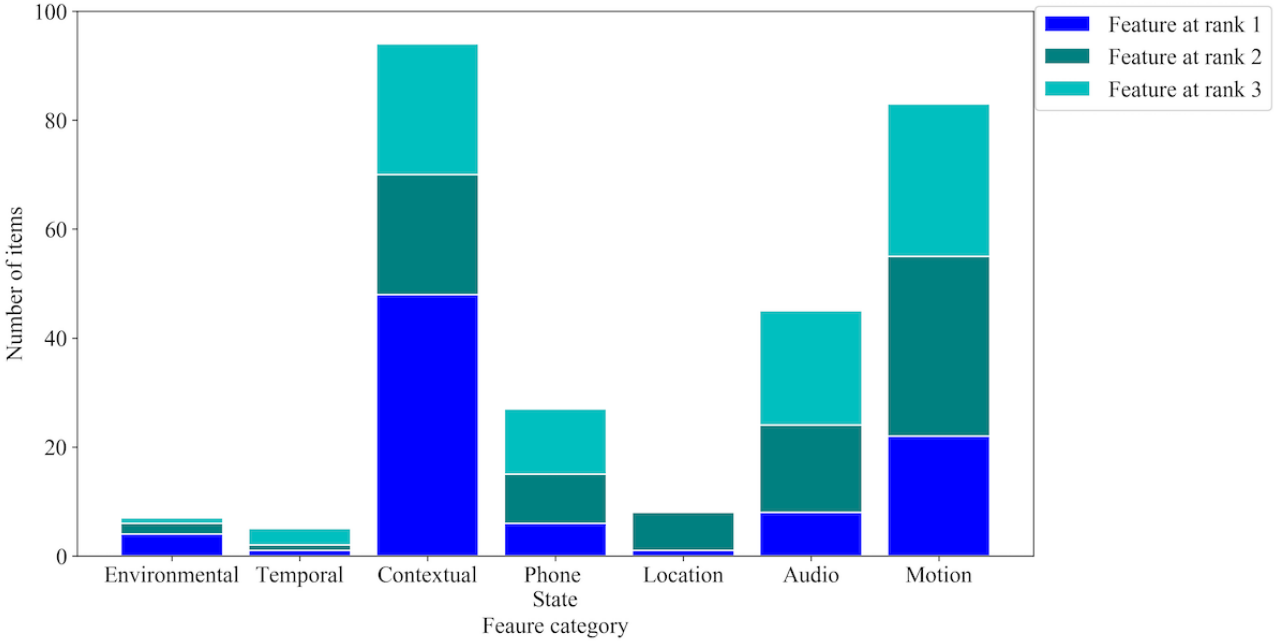


Figure 6. Importance of feature categories for emotional state detection. The color-coded portions of each category represent their contributions to the top 3 ranks of features of 18 persons.

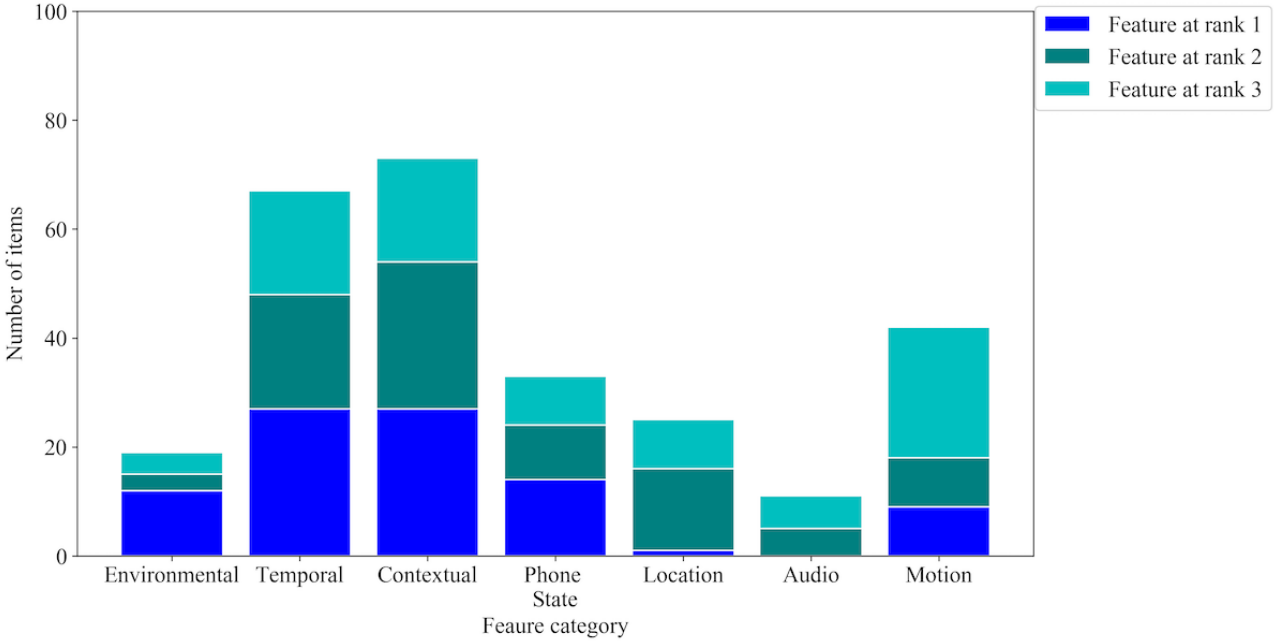


Figure 5 shows that the most important feature categories for emotional transition detection are contextual, motion, and audio signals, whereas the least important categories are temporal and environmental features. In contrast, as shown in Figure 6, the most important feature categories for emotional state detection are contextual and temporal, whereas the least important category is audio signals. Unlike emotional transition, emotional state was more influenced by environmental and location features. This also explains why the personalized emotional state detection models performed better than the general models where the location data were ignored. However, contextual data played the most important role in both emotional transition and state detection.

Discussion

The 4 major findings of this exploratory study are as follows:

- Emotional transitions in small intervals are detectable from data captured via smartphones and smartwatches using ML techniques. We observed that the contextual data, sensed phone states, and motion-related signals are the most influential features for emotional transition detection.
- The prevailing emotional states and the direction in the 3 dimensions of the PAD model are detectable by applying ML algorithms to information captured by smartphones and smartwatches. The contextual and temporal data play

important roles in the detection of emotional states in small time intervals.
- We observed a wide range of interpersonal variations in terms of the detectability of emotional transitions or states. In terms of the personalized models, no single ML model performed the best across all 18 persons.
- Personalized models can better detect emotional states than general models. We believe that a given individual's data contributes most to detecting their own emotional state rather than using data from other individuals.

This study shows the feasibility of detecting persons' emotional states and transitions passively by training ML models on the daily data sensed via smartphones and smartwatches. In practice, these findings will help in reducing self-reports, enabling seamless tracking of daily emotions. For example, a person can be prompted to verify his or her emotional states only if the smartphone app senses a transition in the emotional state of the person. In addition, it can provide the capability of tracing the influential contextual variables that trigger the transition, which can be adjusted by the person for better self-management and well-being. Our investigation showed that self-reported contextual information played the most important role in both emotional transition and state detection. Although the contextual data are self-reported in this data set, latest research shows that contextual data can be labeled from sensed data by smartphone apps automatically without human intervention [28].

The emotional transition detection results obtained in this study are promising but not as good as the emotional state prediction performance. One reason for this could be the highly imbalanced classes, which can be resolved largely by acquiring data for a longer period of time (we had a noticeable improvement in the general models). Moreover, performance can be improved by combining contextual information with other dominant factors of emotions such as personality traits, social communication (eg, incoming-outgoing phone calls, duration, text messages, social media usage), and physiological signals (eg, heart rates, skin conductance) captured via smartphones and smartwatches.

Unlike emotional transition detection, the performance of general models for emotional state detection was poor. Apart from being a harder task (six-class classification), we believe this is also due to the high variability across different individuals. Many state-of-the-art studies stressed the need for further research on interperson variations in affect, mood, and mental health [11,33]. One of the aims of our research was to address this identified gap by exploring subjective variabilities in emotional state and transition detection. Therefore, in addition to the general models, we built personalized models. In these personalized models, each model was trained and tested on each person's data to explore the impact of interpersonal variability on the performance of emotional transition and state detection. The results of the personalized and generic models of this study bolster the need to consider subjective variability while building ML models for emotional state and change detection.

We also observed that simpler models, such as LR, performed better for emotional transition detection and complex models, outperformed during emotional state detection. One reason for this could be the consideration of changes in features between 2 consecutive windows that made the data set sparser (containing many 0s) and smaller for emotional transition detection. Although the personalized models performed better than the general models, especially in the emotional state detection task, the general models can be used as baseline models, which can subsequently be personalized for each person.

Although research has suggested the inclusion of dominance for a better understanding of emotional states [21], it was ignored in the existing works on emotional state and mood recognitions. Our study showed that the prevailing emotional state and its direction in all the 3 dimensions can be detected using ML models on contextual information and data sensed via mobile phones and wearable devices. It can provide data-driven insights on which of the 3 dimensions of emotion prevailed for the person when, where, and in what direction, eventually leading the person toward effective lifestyle changes and better self-management.

Our study shows that a large number of interpersonal variabilities yield superior detection of the emotional transition and state for some persons than others. We manually investigated the reason by considering 3 cases: best, worst, and average, to explore the association between everyday contexts and emotions of individuals. We selected persons 9, 2, and 14 as the best, worst, and average cases, respectively. We plotted the heatmaps of daily spatial contexts, activities, and emotional states over time for the worst, best, and average cases in Figures 7, 8, and 9, respectively. The 3 figures show that the majority of the data were collected when the persons were indoors or at home. For the worst case shown in Figure 7, the day-to-day activities and the emotional states of the person do not exhibit noticeable patterns over time. This might explain why the ML models were unable to capture a strong pattern of the contexts and emotions. On the other hand, for the best case shown in Figure 8, we observe clearer patterns in spatial context, activities, and emotional states over time despite a large amount of missing data. For example, in the best case, the person's emotional state is *dissuaded* while lying down and *pleased* while watching television at home. The average case shown in Figure 9 exhibits some clear patterns of spatial context, activities, and emotional states over time. For example, a person reported *pleasure* mostly while being with friends and *dissuaded* or *discordant* while sleeping at home or outdoors. Therefore, regular patterns in lifestyle are important for the predictability of emotional state and transition detection, and this study showed that such patterns can be captured by leveraging ML algorithms and data acquired via smartphones and consumer-grade wearable devices.

Figure 7. Daily life versus emotional states of person 2 (worst case). The x-axis plots 288 windows per day, and the y-axis plots the number of days in the data collection period of person 2. The color-coded regions represent (a) spatial contexts, (b) activities, and (c) emotional states in each window over the period of data collection (8 days). The white regions represent missing data.

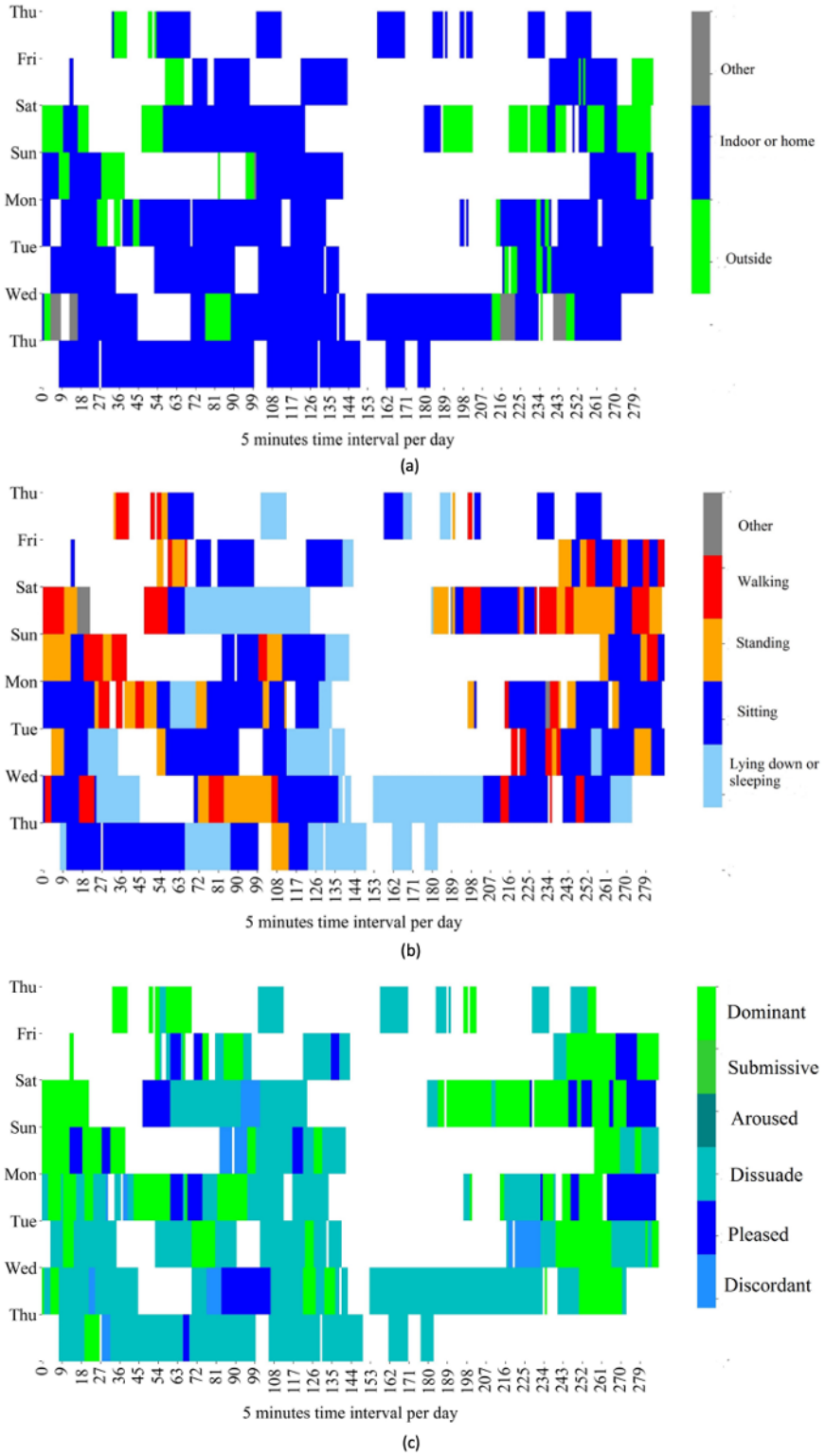


Figure 8. Daily life versus emotional states of person 9 (best case). The x-axis plots 288 windows per day, and the y-axis plots the number of days in the data collection period of person 9. The color-coded regions represent (a) spatial contexts, (b) activities, and (c) emotional states in each window over the period of data collection (8 days). The white regions represent missing data.

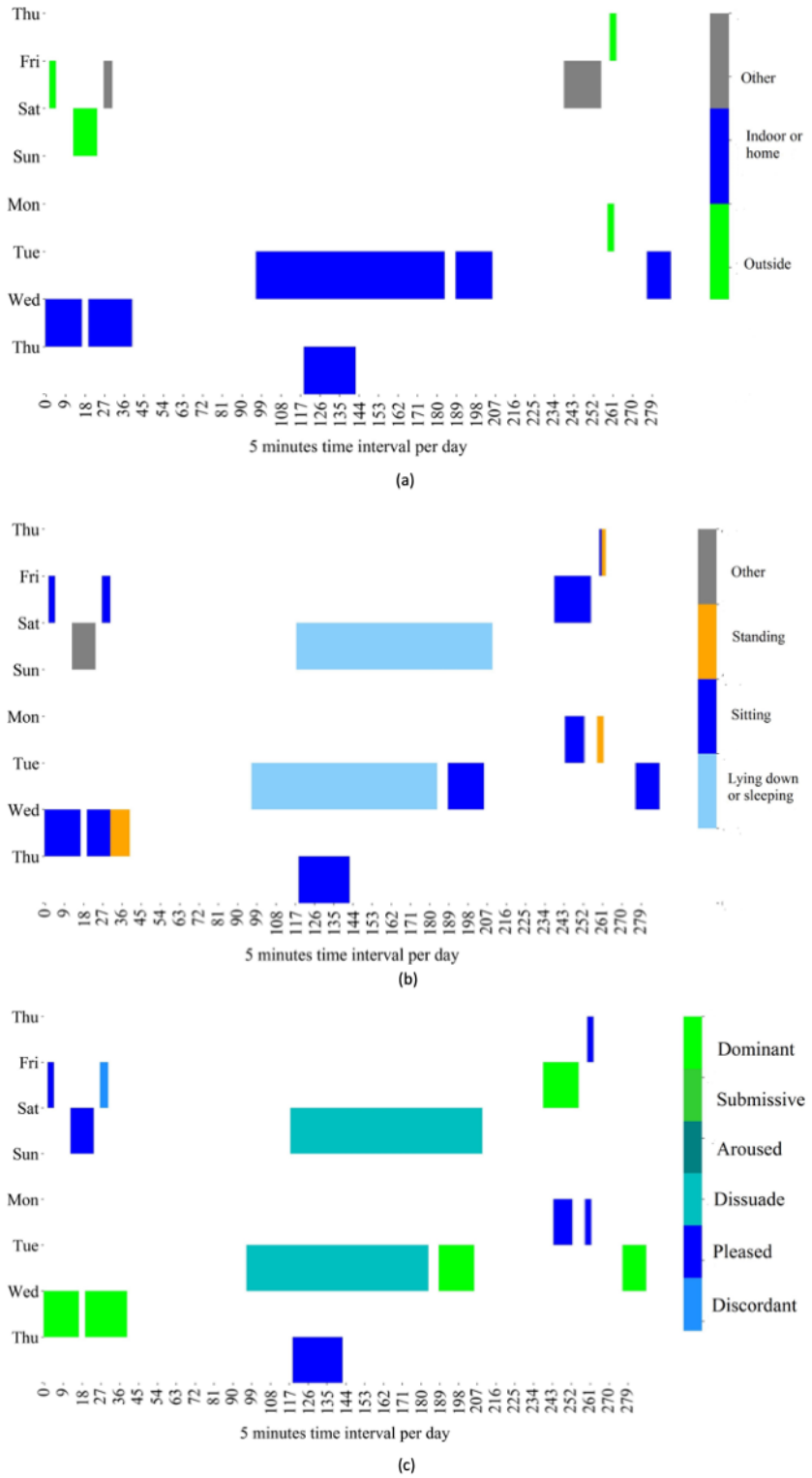
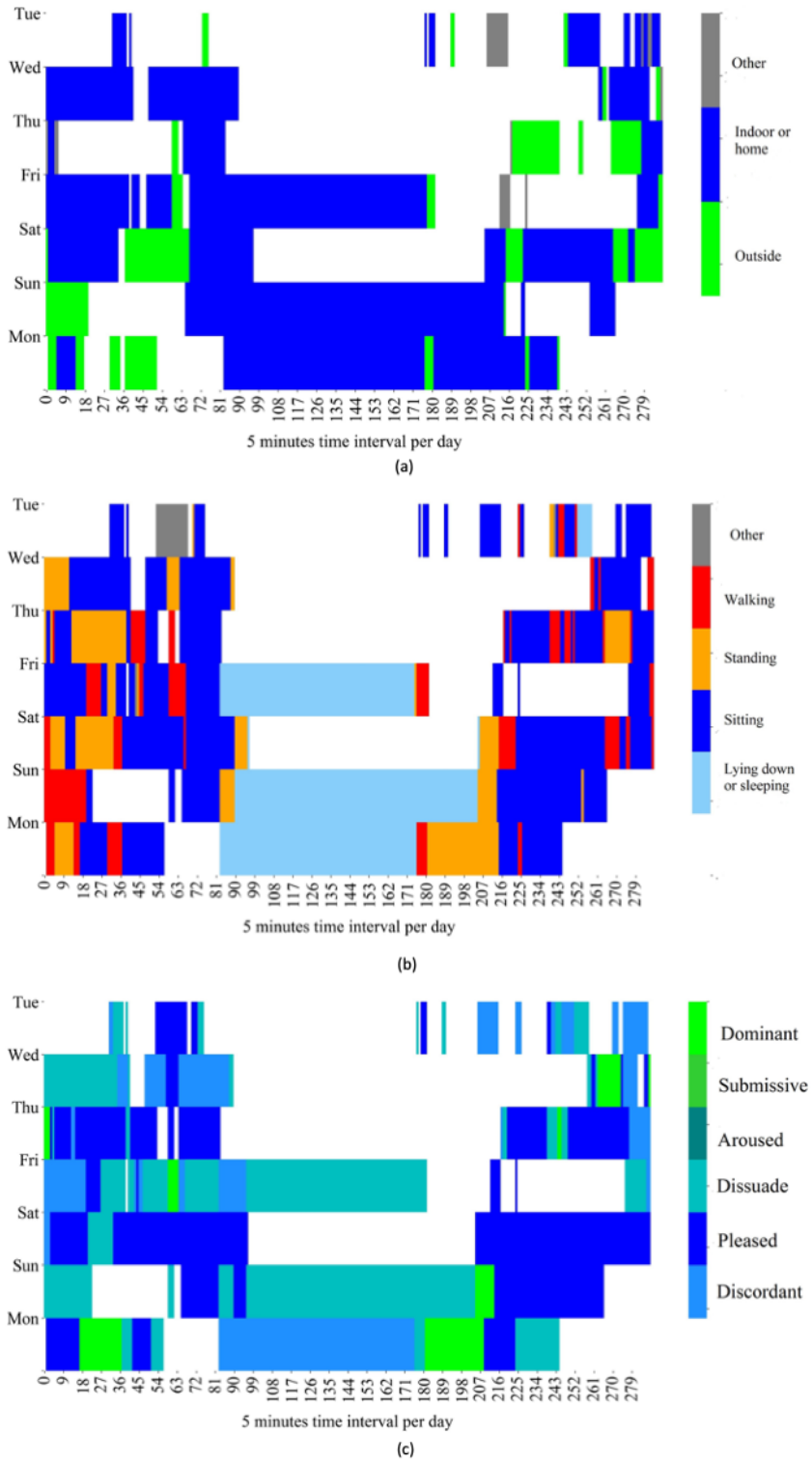


Figure 9. Daily life versus emotional states of person 14 (average case). The x-axis plots 288 windows per day, and the y-axis plots the number of days in the data collection period of person 9. The color-coded regions represent (a) spatial contexts, (b) activities, and (c) emotional states in each window over the period of data collection (7 days). The white regions represent missing data.



Limitations

There are limitations in this study. First, the results are on the basis of a relatively small data set from a small geographical area. In addition, the subjects lacked diversity in age and occupation as the majority were students and researchers at UCSD. Moreover, some spatiotemporal patterns may not have been captured in the data due to large amounts of missing values as well as limited data collection periods (maximum 9 days).

Owing to the small number of participants included in the data set, it was infeasible to hold a test set in conjunction with cross-validation for hyperparameter tuning. As a result, the classification performances reported in this paper are likely overestimated.

In addition, the emotional transition and state classes were highly imbalanced, where 1 or 2 classes of emotional states were absent in some persons' data. Therefore, we suggest replicating our experiments on larger data sets obtained from diverse populations and geographic locations for longer periods of time.

Furthermore, the contextual labels that the classifications were based on were self-reported. Reliance on self-reported data is another limitation of the data set. We assumed that this self-reported information was correct in all cases. Hence, the results from this study partially depend on the accuracy of this self-reported information. To extend this study toward truly passive emotional monitoring, future research should explore eliminating self-reporting by predicting contextual information on the basis of raw sensor data.

Although mapping the emotions into the PAD system results in 3 dimensions, we focused only on the most dominant one. This simplified the rich, multidimensional information that the PAD system provided for making classification tasks more feasible. An alternative is to create a new emotion mapping system specifically for the ExtraSensory data set via clustering in the 3D space defined by PAD. This is a worthwhile future research direction.

Another major limitation is the absence of health-related information such as BMI, gender, age, and mental health biomarkers in the data. Therefore, further investigation is needed to shed light on the association between health status and emotional transition and state of persons.

Conclusions

In this study, we explored the feasibility of detecting emotional transitions and states by applying ML techniques to daily data captured via smartphones and smartwatches. Our results established an association between emotional transition and state and contextual information. We also investigated the salient contextual variables influencing emotional states and transitions. The interpersonal variability in our results bolsters the need for further research on personalized prediction of emotional states and transitions. The findings of this study support the utility of passive data collection, reduced self-reporting, enhanced tracking of psychological well-being, self-awareness, self-management, and risk prediction and just-in-time interventions.

Acknowledgments

This work was supported by a Postdoctoral Scholarship from the O'Brien Institute for Public Health and Research Funding from the Libin Cardiovascular Institute of Alberta at the University of Calgary.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Dataset description and extended experimental results.
[DOCX File , 60 KB - mhealth_v8i9e17818_app1.docx]

References

1. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr DC. The relationship between clinical, momentary, and sensor-based assessment of depression. Int Conf Pervasive Comput Technol Healthc 2015 Aug;2015:- [FREE Full text] [doi: 10.4108/icst.pervasivehealth.2015.259034] [Medline: 26640739]
2. Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan;36(1):72-81. [doi: 10.1002/da.22822] [Medline: 30129691]
3. Servia-Rodríguez S, Rachuri K, Mascolo C. Mobile Sensing at the Service of Mental Well-being: A Large-scale Longitudinal Study. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia. [doi: 10.1145/3038912.3052618]
4. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019;2:3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
5. Torous J, Staples P, Onnela J. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602 [FREE Full text] [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
6. Difrancesco DS, Fraccaro P, Veer SN. Out-of-Home Activity Recognition from GPS Data in Schizophrenic Patients. In: 29th International Symposium on Computer-Based Medical Systems. 2016 Presented at: CBMS'16; June 20-24, 2016; Dublin, Ireland. [doi: 10.1109/cbms.2016.54]
7. Osmani V. Smartphones in mental health: detecting depressive and manic episodes. IEEE Pervasive Comput 2015 Jul;14(3):10-13. [doi: 10.1109/MPRV.2015.54]
8. Suhara Y, Xu Y, Pentland A. DeepMood: Forecasting Depressed Mood Based on Self-Reported Histories via Recurrent Neural Networks. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia. [doi: 10.1145/3038912.3052676]

9. Berrouiguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Carmona Camacho R, Baca-García E, et al. Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: a case series of the evidence-based behavior (EB2) study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e197 [FREE Full text] [doi: 10.2196/mhealth.9472] [Medline: 30530465]
10. Helbich M. Toward dynamic urban environmental exposure assessments in mental health research. Environ Res 2018 Feb;161:129-135 [FREE Full text] [doi: 10.1016/j.envres.2017.11.006] [Medline: 29136521]
11. Sandstrom GM, Lathia N, Mascolo C, Rentfrow PJ. Putting mood in context: using smartphones to examine how people feel in different locations. J Res Pers 2017 Aug;69:96-101. [doi: 10.1016/j.jrp.2016.06.004]
12. Ma Y, Xu B, Bai Y. Daily Mood Assessment Based on Mobile Phone Sensing. In: Ninth International Conference on Wearable and Implantable Body Sensor Networks. 2012 Presented at: BSN'12; May 9-12, 2012; London, UK. [doi: 10.1109/bsn.2012.3]
13. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
14. Wang R, Chen F, Chen Z. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp'14; September 13-17, 2014; Seattle, USA. [doi: 10.1145/2632048.2632054]
15. Sano A, Phillips A, Yu A. Recognizing Academic Performance, Sleep Quality, Stress Level, and Mental Health Using Personality Traits, Wearable Sensors and Mobile Phones. In: 12th International Conference on Wearable and Implantable Body Sensor Networks. 2015 Presented at: BSN'15; June 9-12, 2015; Cambridge, MA, USA. [doi: 10.1109/bsn.2015.7299420]
16. Budner P, Eirich J, Gloor P. 'Making you happy makes me happy' - measuring individual mood with smartwatches. arXiv 2017 epub ahead of print [FREE Full text]
17. Zhang X, Li W, Chen X, Lu S. MoodExplorer: Towards Compound Emotion Detection via Smartphone Sensing. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018 Presented at: ACM'18; September 9-12, 2018; London, UK. [doi: 10.1145/3161414]
18. Umematsu T, Sano A, Picard R. Daytime Data and LSTM can Forecast Tomorrow's Stress, Health, and Happiness. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2019 Presented at: EMBC'19; July 23-27, 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8856862]
19. Spathis D, Servia-Rodriguez S, Farrahi K. Sequence Multi-task Learning to Forecast Mental Wellbeing from Sparse Self-Reported Data. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Presented at: KDD'19; August 4-8, 2019; Anchorage, AK, USA. [doi: 10.1145/3292500.3330730]
20. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
21. Bakker I, van der Voordt T, Vink P, de Boon J. Pleasure, arousal, dominance: Mehrabian and Russell revisited. Curr Psychol 2014 Jun 11;33(3):405-421. [doi: 10.1007/s12144-014-9219-4]
22. The ExtraSensory Dataset. Yonatan Vaizman - University of California San Diego. URL: http://extrasensory.ucsd.edu/ [accessed 2020-08-24]
23. Vaizman Y, Ellis K, Lanckriet G. Recognizing detailed human context in the wild from smartphones and smartwatches. IEEE Pervasive Comput 2017 Oct;16(4):62-74. [doi: 10.1109/mprv.2017.3971131]
24. Mehrabian A, Russell J. An Approach to Environmental Psychology. New York, USA: MIT Press; 1974.
25. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J Res Pers 1977 Sep;11(3):273-294. [doi: 10.1016/0092-6566(77)90037-x]
26. Bradley MM, Lang PJ. Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. Semantic Scholar. 1999. URL: https://www.uvm.edu/pdodds/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf [accessed 2020-09-01]
27. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res Methods 2013 Dec;45(4):1191-1207. [doi: 10.3758/s13428-012-0314-x] [Medline: 23404613]
28. Vaizman Y, Ellis K, Lanckriet G. ExtraSensory App: Data Collection In-the-Wild with Rich User Interface to Self-Report Behavior. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 21-26, 2018; Montreal, QC, Canada. [doi: 10.1145/3173574.3174128]
29. Geohash. URL: http://geohash.org [accessed 2020-08-25]
30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002 Jun 1;16:321-357. [doi: 10.1613/jair.953]
31. mHealth Emotional State-Transition Detection Public. GitHub. URL: https://github.com/data-intelligence-for-health-lab/mHealth_Emotional_State-Transition_Detection_Public [accessed 2020-08-25]
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. New York, USA: MIT Press; 2017.
33. Jaques N, Taylor S, Sano A. Predicting tomorrow's mood, health, and stress level Using personalized multitask learning and Domain adaptation. J Mach Learn Res 2017;66:17-33 [FREE Full text]

Abbreviations

ANEW: Affective Norms for English Words
AUROC: area under the receiver operating characteristic
CB: CatBoost
LR: logistic regression
ML: machine learning
PAD: pleasure, arousal, dominance
RF: random forest
SHAP: shapley additive explanation
SVMSMOTE: Support Vector Machines Synthetic Minority Over-sampling Technique
UCSD: University of California, San Diego
XGB: XGBoost

Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by Y Vaizman, S Tedesco; comments to author 10.03.20; revised version received 09.07.20; accepted 03.08.20; published 29.09.20.

Please cite as:
Sultana M, Al-Jefri M, Lee J
Using Machine Learning and Smartphone and Smartwatch Data to Detect Emotional States and Transitions: Exploratory Study
JMIR Mhealth Uhealth 2020;8(9):e17818
URL: http://mhealth.jmir.org/2020/9/e17818/
doi:10.2196/17818
PMID:32990638

Original Paper

Daily Activities Related to Mobile Cognitive Performance in Middle-Aged and Older Adults: An Ecological Momentary Cognitive Assessment Study

Laura M Campbell[1,2], MS; Emily W Paolillo[1,2], MS; Anne Heaton[2], BA; Bin Tang[2], PhD; Colin A Depp[2], PhD; Eric Granholm[2,3], PhD; Robert K Heaton[2], PhD; Joel Swendsen[4,5], PhD; David J Moore[2], PhD; Raeanne C Moore[2], PhD

[1]SDSU/UC San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, United States

[2]Department of Psychiatry, University of California San Diego, San Diego, CA, United States

[3]VA San Diego Healthcare System, San Diego, CA, United States

[4]CNRS UMR, University of Bordeaux, Bordeaux, France

[5]National Center for Scientific Research, Ecole Pratique des Hautes Etudes, PSL Research University, Paris, France

Corresponding Author:
Raeanne C Moore, PhD
Department of Psychiatry
University of California San Diego
220 Dickinson St. Suite B (8231)
San Diego, CA, 92103-8231
United States
Phone: 1 6195435378
Email: r6moore@ucsd.edu

Abstract

Background: Daily activities have been associated with neurocognitive performance. However, much of this research has used in-person neuropsychological testing that requires participants to travel to a laboratory or clinic, which may not always be feasible and does not allow for the examination of real-time relationships between cognition and behavior. Thus, there is a need to understand the real-time relationship between activities in the real world and neurocognitive functioning to improve tracking of symptoms or disease states and aid in the early identification of neurocognitive deficits among at-risk individuals.

Objective: We used a smartphone-based ecological momentary cognitive assessment (EMCA) platform to examine real-time relationships between daily activities and neurocognitive performance (executive functioning and verbal learning) in the everyday environment of middle-aged and older adults with and without HIV.

Methods: A total of 103 adults aged 50-74 years (67 persons with HIV; mean age 59 years, SD 6.4) were recruited from the University of California, San Diego HIV Neurobehavioral Research Program and the San Diego community. Participants completed our EMCA protocol for 14 days. Participants reported their current daily activities 4 times per day; following 2 of the 4 daily ecological momentary assessment (EMA) surveys, participants were administered the mobile Color-Word Interference Test (mCWIT) and mobile Verbal Learning Test (mVLT), each once per day. Activities were categorized into cognitively stimulating activities, passive leisure activities, and instrumental activities of daily living (IADLs). We used multilevel modeling to examine the same-survey and lagged within-person and between-person effects of each activity type on mobile cognitive performance.

Results: On average, participants completed 91% of the EMA surveys, 85% of the mCWIT trials, and 80% of the mVLT trials, and they reported engaging in cognitively stimulating activities on 17% of surveys, passive leisure activities on 33% of surveys, and IADLs on 20% of surveys. Adherence and activity percentages did not differ by HIV status. Within-persons, engagement in cognitively stimulating activities was associated with better mCWIT performance (β=−1.12; *P*=.007), whereas engagement in passive leisure activities was associated with worse mCWIT performance (β=.94; *P*=.005). There were no lagged associations. At the aggregate between-person level, a greater percentage of time spent in cognitively stimulating activities was associated with better mean mVLT performance (β=.07; *P*=.02), whereas a greater percentage of time spent in passive leisure activities was associated with worse mean mVLT performance (β=−.07; *P*=.01). IADLs were not associated with mCWIT or mVLT performance.

Conclusions: Smartphones present unique opportunities for assessing neurocognitive performance and behavior in middle-aged and older adults' own environment. Measurement of cognition and daily functioning outside of clinical settings may generate

novel insights on the dynamic association of daily behaviors and neurocognitive performance and may add new dimensions to understanding the complexity of human behavior.

(JMIR Mhealth Uhealth 2020;8(9):e19579) doi:10.2196/19579

KEYWORDS

ecological momentary assessment; daily functioning; telemedicine; digital health; neuropsychological test; cognition; HIV; aging; mobile phone

Introduction

Background

Traditional neuropsychological testing provides a *snapshot* of a patient's neurocognitive functioning at one time point in an optimal, controlled environment (ie, without distractions). This assessment method allows neuropsychologists to make empirically based judgments about a number of critical patient factors (eg, neurocognitive impairment status and likely etiology of deficits). However, there are several limitations associated with traditional laboratory-based neuropsychological assessments. First, traditional testing requires in-person, face-to-face contact. The limitations of this are more salient now than ever before owing to the need for social distancing because of the COVID-19 pandemic [1]. However, even before the COVID-19 pandemic, there were several barriers associated with in-person neuropsychological testing. For example, it is difficult for individuals with limited access to transportation or those who live in rural areas to travel to clinics and/or participate in research. Second, although one goal of neuropsychological assessment is often to make judgments about everyday functioning, the connection between performance on neuropsychological testing and neurocognitive functioning in daily life is imperfect [2]—people do not always perform consistently in an optimal manner in everyday life. In other words, a person's neurocognitive capacity, as demonstrated in a neuropsychological testing room, may not predict their performance in everyday tasks that take place in a more distracting, real-world environment [3,4].

Now, more than ever, we need to leverage digital health tools to improve upon assessment of cognition [5,6]. One feasible solution to the limitations of traditional neuropsychological testing is the utilization of mobile cognitive testing. This burgeoning digital health assessment method involves objective cognitive tests that are self-administered through a smartphone, tablet, or other mobile device. Thus, mobile cognitive testing both reduces the need for in-person visits and allows for repeated assessment of neurocognitive functioning in an individual's natural environment [7,8]. Therefore, this methodology may be particularly useful for detecting neurocognitive decline earlier, given that performance on tests can be easily tracked over time and repeatedly compared with one's own previous performance. Moreover, mobile cognitive testing may be more ecologically valid and thus help to better understand subtle changes in neurocognitive functioning as they occur in participants' own environments. Despite some technological challenges (eg, standardizing stimuli and response latencies across different software and hardware platforms) [9], mobile cognitive testing has been found to be feasible and valid among various clinical [7,10,11] and nonclinical populations [7,10-12].

Mobile cognitive testing may be a particularly important tool for a better understanding of the relationship between real-world neurocognitive and everyday functioning. In the context of a clinical neuropsychological evaluation, the relationship between cognition and everyday functioning is often considered to be unidirectional, as neuropsychologists use neurocognitive test data to predict everyday functioning outcomes. However, a bidirectional relationship between certain everyday activities and neurocognitive functioning has been demonstrated in older adults [13]. For example, time spent in passive leisure activities (eg, watching television) has been shown to be inversely related to cross-sectional neurocognitive performance and positively related to the likelihood of a neurocognitive decline in subsequent years [14-18]. Conversely, studies have also demonstrated a positive relationship between time spent in cognitively, physically, and/or socially stimulating activities and neurocognitive functioning in older adults [19] and persons with HIV [20]. The existing literature examining relationships between everyday activities and cognition, however, is limited in several ways. Self-report retrospective measures may be inaccurate because of recall error and response biases [21,22], whereas performance-based measures of everyday functioning are limited in their ability to be administered frequently (eg, need for in-person clinic visits, practice effects) and were developed to measure functional capacity, that is, whether a person has the capacity to function independently. However, many are not validated as tools to assess functional performance, that is, how people actually function in their home environments [23]. In addition, most studies have only examined between-person effects (eg, baseline activities predicting change in cognition over time) [17]. Thus, little is known about the possible acute and dynamic relationships between everyday activities and cognition within individuals.

Coupling mobile cognitive testing with ecological momentary assessment (EMA; ie, repeated self-report assessment of in-the-moment feelings, behaviors, and contexts) affords the opportunity to examine real-world, real-time relationships between daily activities and neurocognitive performance. For example, in the only study to date (to our knowledge) that has examined the within-person relationship between daily activities and cognition using EMA and mobile cognitive testing, Allard et al [24] found that engagement in cognitively stimulating activities was associated with better semantic memory performance later in the day in a sample of older adults. However, participation in passive leisure activities was not significantly associated with any differences in performance on the mobile test of semantic memory [24]. This study serves as

an example of how understanding the dynamic relationships between daily life activities and cognition could lead to improved symptoms or disease tracking or just-in-time adaptive interventions to promote optimal neurocognitive functioning, when desired, using digital health technologies. Such interventions may be important for older adults [25] and persons with HIV [26,27], particularly older persons with HIV, given the higher rates of neurocognitive impairment compared with the general population.

Objectives

In light of these questions, this study used a smartphone-based ecological momentary cognitive assessment (EMCA) platform (ie, a custom-built integrated EMA and mobile cognitive testing platform) to examine the real-world relationships between daily activities (ie, cognitively stimulating activities, passive leisure activities, and instrumental activities of daily living [IADLs]) and neurocognitive performance (ie, executive functioning and learning) among older persons with and without HIV. The first aim of this study is to examine (1) the same-survey, within-person relationships between reported activities and mobile cognitive performance and (2) the between-person effect of the percentage of surveys in which the activity was endorsed with mean mobile cognitive test performance. Accounting for between-person effects allows for differentiation between whether an activity is associated with true changes and/or fluctuations in cognition or whether more total time spent in an activity, in general, tends to be associated with a higher or lower level of cognition. The second aim is to examine the temporal ordering of effects, with activity engaged in at one survey predicting cognition at the same-day next survey, administered approximately 3 to 4 hours later (ie, lagged analyses). We hypothesized that in all analyses, cognitively stimulating activities would be associated with better mobile cognitive performance, and passive leisure activities would be associated with worse mobile cognitive performance. In addition, we hypothesized that engagement in IADLs would be associated with better mobile cognitive performance, but the effect would be weaker than that observed with cognitively stimulating activities. Given that potential differences in observed effects by HIV serostatus would be likely because of differences in a combination of sociodemographic and/or environmental factors rather than HIV itself, we accounted for HIV status but did not specifically examine any HIV interactions.

Methods

Participants

A total of 67 community-dwelling persons with HIV and 36 HIV-negative middle-aged and older adults, aged 50 to 74 years, were included in this National Institute of Mental Health (NIMH)–funded study at the HIV Neurobehavioral Research Program (HNRP) at the University of California, San Diego (UCSD) between 2016 and 2019. Participants were recruited from the participant pool at the HNRP or through the community (eg, HIV clinics, flyers, and community centers). Inclusion criteria for the study were being aged ≥50 years, the ability to provide written informed consent, and being fluent in English. Exclusion criteria for the study were self-reported histories of serious mental illness (eg, schizophrenia and bipolar disorder), non-HIV neurological disorder (eg, stroke), head injury with loss of consciousness for ≥30 min, or indication of a severe learning disability (as indicated by a score of <70 on the Wide Range Achievement Test, fourth edition Reading Subtest [WRAT-4]). Participants with a positive urine toxicology test (with the exception of marijuana) or alcohol breathalyzer on the day of the in-person visit were rescheduled. All procedures were approved by UCSD's Institutional Review Board before protocol implementation, and all participants demonstrated decisional capacity [28] and provided written informed consent.

Measures and Procedures

Laboratory Visits

Participants completed an initial in-person baseline visit that included a tutorial on the EMCA portion of the study and a comprehensive neuromedical and neurobehavioral assessment. Participants were given a Samsung Galaxy S 4.2 YP-GI1 8GB smartphone with a 4G Android operating system (OS) for the duration of the study. The Galaxy Player 4.2 has a 4.2“ IPS (in-plane switching) display at 800×480 resolution, 1 GHz processor, using Android 2.3.6 Gingerbread OS. Participants were provided an individualized, face-to-face 20- to 30-min tutorial with a research associate on how to complete EMA surveys and mobile cognitive tests and were given a smartphone operating manual to take home.

Psychiatric and substance use disorders in Table 1 were determined via a computer-assisted structured interview, that is, the Composite International Diagnostic Interview [29]. In-person neurocognitive functioning in Table 1 was determined by comprehensive neuropsychological assessment (previously described in detail by Heaton et al [27]; see Multimedia Appendix 1 for specific neuropsychological tests), and neurocognitive domain scaled scores (mean 10, SD 3) that adjust for practice effects were generated [30]. Impairment was determined using a global deficit score [31] of ≥0.5, and premorbid verbal IQ was measured via the WRAT-4 [32]. For all participants, HIV serostatus was confirmed with HIV/Hepatitis C virus antibody point-of-care rapid test (Miriad-MedMira) and confirmed by western blot analyses. Among persons with HIV, AIDS diagnosis, antiretroviral therapy (ART) regimen, estimated duration of HIV disease, and nadir CD4 count were obtained by self-report (unless the current CD4 count was lower than the reported nadir CD4 value). Viral load detectability (≥50 copies/mL) and the current CD4 count was measured in blood plasma.

At the end of the 14-day EMA study period, participants returned to the HNRP to deliver the study phone and completed additional neuropsychological tests and study questionnaires. Participants were compensated for both study visits. Bonus compensation (US $1 per survey) was provided for each EMA survey participants completed.

Table 1. Participants' characteristics by HIV serostatus (N=103).

Participant characteristics	HIV+ (n=67)	HIV− (n=36)	*t* value or chi-square	*P* value
Demographic variables				
Age (years), mean (SD)	59.3 (6.3)	59.2 (6.7)	0.0	.99
Sex (male), n (%)	53 (79)	20 (56)	6.3	.01
Race or ethnicity, n (%)	N/A[a]	N/A	1.3	.73
Non-Hispanic White	42 (63)	23 (64)		
African American	16 (24)	6 (17)		
Hispanic	4 (10)	6 (17)		
Other	2 (3)	1 (3)		
Education (years), mean (SD)	13.9 (2.4)	15.0 (2.5)	−2.1	.04
Employed[b], n (%)	20 (31)	14 (40)	0.9	.35
Household income[c] (US$), n (%)	N/A	N/A	FET[d]	.002
<10,000	12 (18)	6 (17)		
10,000-34,999	45 (68)	13 (36)		
35,000-74,999	4 (6)	8 (22)		
≥75,000	5 (8)	9 (25)		
Live alone, n (%)	38 (57)	11 (31)	6.6	.01
Smartphone ownership, n (%)	60 (90)	30 (83)	0.8	.36
Psychiatric comorbidities, n (%)				
LT[e] MDD[f]	48 (72)	9 (25)	20.6	<.001
Current MDD[g,h]	11 (16)	1 (3)	FET	.05
LT any substance use disorder	45 (67)	17 (47)	3.9	.04
Current substance use disorder[h,i]	2 (3)	1 (3)	FET	.99
In-person neurocognitive functioning				
GDS[j] impaired[c], n (%)	18 (27)	7 (20)	0.59	.44
Global SS[c,k], mean (SD)	9.2 (2.0)	9.9 (1.8)	−1.6	.11
Verbal SS[c], mean (SD)	10.2 (2.7)	11.4 (2.8)	−2.0	.05
Executive functioning SS[c], mean (SD)	8.9 (2.5)	9.8 (2.0)	−1.7	.10
Speed of information processing SS[c], mean (SD)	9.6 (2.4)	10.4 (2.5)	−1.4	.16
Learning SS[c], mean (SD)	8.3 (2.3)	9.1 (2.3)	−1.8	.07
Recall SS[c], mean (SD)	8.6 (2.2)	9.3 (2.4)	−1.5	.14
Working memory SS[c], mean (SD)	9.8 (2.9)	10.3 (2.8)	−0.8	.42
Motor SS[c], mean (SD)	7.8 (2.7)	8.0 (2.4)	−0.4	.69
Premorbid verbal IQ[h,l], mean (SD)	102 (14)	106 (16)	−1.2	.24
HIV characteristics				
AIDS[m], n (%)	46 (70)	N/A	N/A	N/A
Current CD4[m], median (IQR)	703 (550-893)	N/A	N/A	N/A
Nadir CD4[m], median (IQR)	148 (33-285)	N/A	N/A	N/A
Duration of HIV infection (years), median (IQR)	23.8 (15.7-28.8)	N/A	N/A	N/A

Participant characteristics	HIV+ (n=67)	HIV− (n=36)	*t* value or chi-square	*P* value
On antiretroviral therapy, n (%)	63 (94)	N/A	N/A	N/A
Undetectable viral load[n], n (%)	60 (97)	N/A	N/A	N/A

[a]N/A: not applicable.

[b]n=100.

[c]n=102.

[d]FET: Fisher exact test.

[e]LT: lifetime.

[f]LT MDD: met criteria for major depressive disorder at any point in life.

[g]Current MDD: currently meets criteria for major depressive disorder.

[h]n=101.

[i]All current substance use disorders were marijuana use disorder.

[j]GDS: global deficit score.

[k]SS: scaled score; scaled score is based on comprehensive in-laboratory neuropsychological testing (scale score: mean 10, SD 3).

[l]Premorbid Verbal IQ was estimated using the Wide Range Achievement Test, fourth edition Reading Subtest.

[m]n=66.

[n]n=62.

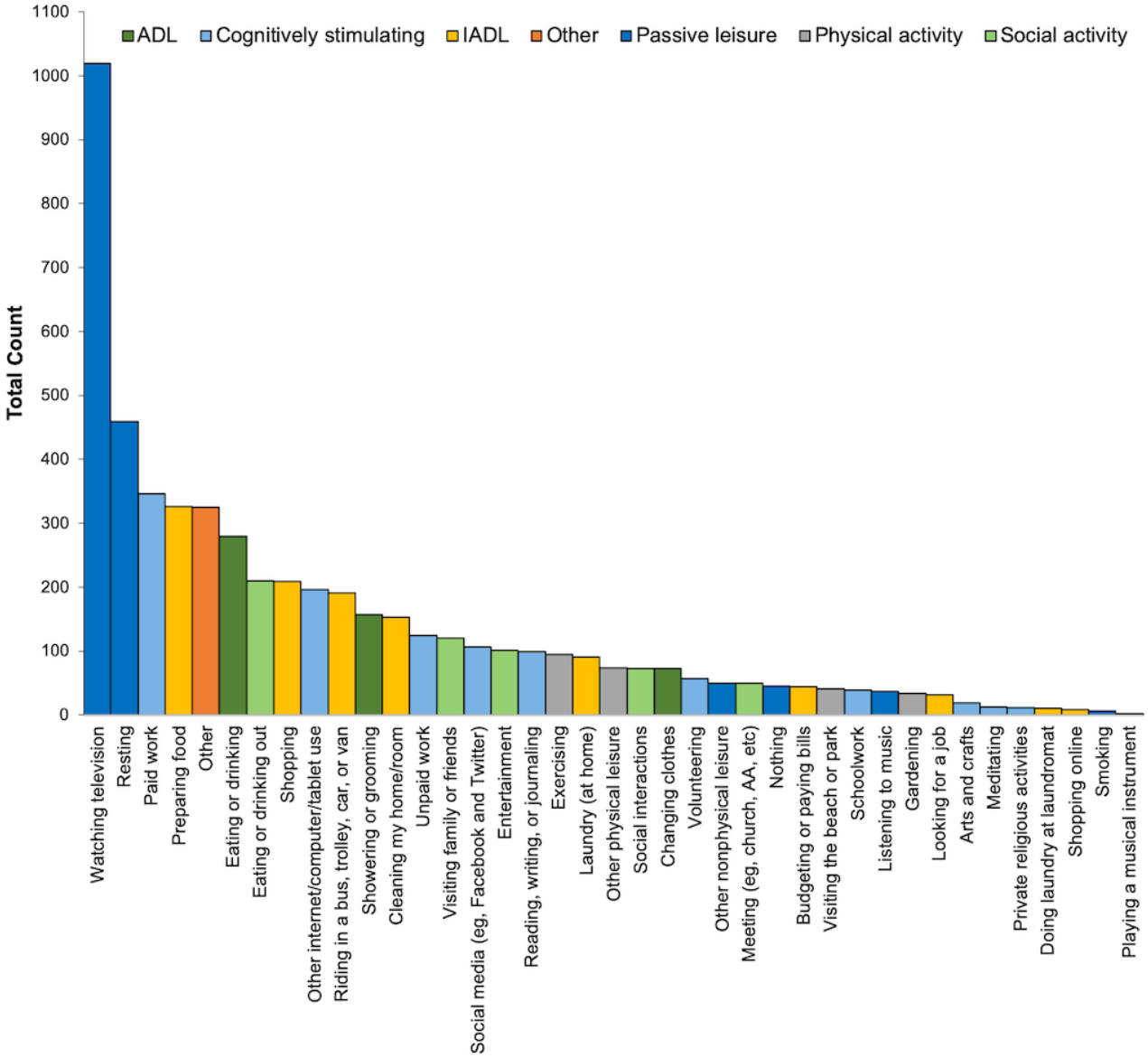
Fourteen-Day EMA and Mobile Cognitive Testing Monitoring

Following their first in-person visit, participants completed the 14-day EMCA protocol. Participants received 4 EMA surveys per day, which occurred at pseudorandom times throughout the day (ie, spaced for a survey to occur in the morning, midday, afternoon, and evening) approximately 3 to 4 hours apart according to each participant's sleep-wake schedule. The study phone alert sounded to signal participants to complete each survey. Once the alert sounded, participants had 10 min to start the survey, with a reminder alarm every 2 min during that period before the survey was considered missed. Participants also had the option to cancel the survey during the 10-min window or at any point during the survey. If a participant only completed part of the survey and then canceled, the data completed were saved. At the end of 2 of the 4 daily surveys, participants were prompted to complete either a mobile cognitive test of executive functioning or verbal learning. The surveys after which participants received the mobile cognitive tests were randomized to different surveys per day and were presented at the same survey per day for each participant. The mobile Color-Word Interference Test (mCWIT) and mobile Verbal Learning Test (mVLT) were not given on the same survey. The study phone's OS was encrypted, in the event that the phone was lost or stolen, to safeguard participants' data. Furthermore, the study phones were locked so that the survey program was the only program on the phone that could be used. Participants were provided contact information in the event that they experienced technological difficulties and were called twice during the 14-day period to assess if they had any difficulties.

Daily Activities

At each survey, participants were asked to report their current activity (ie, "What are you doing?"). Participants could choose 1 response from 39 options that included different activities (Figure 1), with different options available, depending on whether the participant reported being at home or not at home on a previous question. Participants were only given the option to report 1 activity in response to this item and were instructed to choose the primary activity in which they were currently participating. Activities were then categorized into cognitively stimulating activities, passive leisure activities, IADLs, activities of daily living (ADLs), physical activity, social activities, and other activities (ie, if participants endorsed *other* as their activity). Cognitively stimulating activities included working (paid or unpaid), volunteering, schoolwork, arts and crafts, meditating, playing a musical instrument, private religious activities, reading or writing or journaling, and *other internet or computer or tablet use*. Passive leisure activities included: watching television, listening to music, other nonphysical leisure, resting, smoking, social media, and nothing. IADLs included budgeting or paying bills, cleaning, doing laundry, looking for a job, preparing food, traveling (ie, riding in a bus, trolley, car, or van), and shopping. These activities were categorized by 3 authors (RM, CD, and EG) with expertise in this area of research and who reviewed the current literature. ADLs, physical activity, social activities, and *other* were each endorsed on less than 10% of surveys and were thus not examined in this study. Cognitively stimulating activities, passive leisure activities, and IADLs were dichotomized into 0 (*not doing activity*) and 1 (*doing activity*). The percentage of time spent in each activity was calculated by dividing the number of surveys in which an activity was endorsed by the total number of completed EMA surveys over the 14-day study period.

Figure 1. Count of reported activities from the ecological momentary assessment question "What are you doing?" Participants could only report one activity and were instructed to choose the primary activity in which they were currently participating. Participants could choose "other" if their current activity was not listed as an option. AA: alcoholics anonymous; ADL: activities of daily living; IADL: instrumental activities of daily living.



Mood

Depressed mood was assessed at each EMA survey, and response choices for the prompt, "I feel depressed...," were on a 5-point scale ranging from *1=not at all* to *5=very much*. Although not the main focus of this study, the current depression rating was included as a covariate in follow-up analyses to account for mood as a possible confounding factor.

Mobile Cognitive Tests

Participants completed an mCWIT and mVLT once per day. The survey in which they were prompted to complete the mobile cognitive test was randomized and counterbalanced across the different times of the day for the 14-day EMA period. The mCWIT is a test of executive functioning based on the Stroop paradigm whereby people are asked to name the color of the ink in which a color-word is written when the ink color and word are incongruent. Stroop tasks have been used widely among persons with HIV [33,34]; however, there is only one other group to our knowledge that has developed a mobile Stroop-like test for use in an EMA study [11]. Participants were instructed, "Do not read the words, say the colors in which they are written" and had up to 60 seconds to complete the task. Responses were audio-recorded and scored by 2 independent raters. All discrepant scores were scored by a third rater. The outcome assessed in this study was completion time (seconds). mCWIT data were excluded from analyses for 2 participants; one participant was excluded due to colorblindness, and the other participant was excluded because they had an average of 15 of 16 errors, and therefore their data were not considered valid. The reliability and validity of the mCWIT have been described in detail elsewhere [10].

The mVLT is a test designed to assess verbal learning and recall. Although there are other groups who have developed verbal learning or memory tests for use in mobile cognitive testing studies, none have examined these among persons with HIV [8]. Participants were presented with a list of 12 semantically

unrelated words to read on the smartphone for three 30-second trials. A unique list was presented each day. After each trial, participants were asked to recall words from the list and then to select *done* when they finished recalling words. Responses were audio-recorded and scored for the total number of correct responses. The outcome assessed in this study was the total number of words correctly recalled over the 3 learning trials. The mVLT was scored by 2 independent raters, and discrepant scores were reviewed by an additional rater. The association between the mVLT (mean average across 14 days) and in-laboratory Hopkins Verbal Learning Test-Revised [35] was Cohen *d*=1.09. On both tests, there was some intraindividual variability in performance across days, as expected (average SD per person mCWIT=5.2 seconds; mVLT=3.5 words). mCWIT and mVLT trials were excluded from this study if raters suspected cheating (eg, help from others), if the participant was doing something else during the test (eg, talking with others), or if the participant endorsed the use of illicit substances (ie, cocaine, methamphetamine, ecstasy, heroin, *other street drug(s)*, or "prescription drugs not prescribed to me") on any survey before taking the mobile cognitive test (within the same day). In follow-up analyses, analyses were rerun excluding surveys in which participants endorsed alcohol or marijuana use on the same survey.

Statistical Analyses

Participant characteristics by HIV serostatus are presented in Tables 1 and 2 and were assessed via two-tailed chi-square, Fisher exact tests, and *t* tests (or nonparametric equivalent), as appropriate. Separate linear mixed-effects regressions with participant-specific random intercepts were used to evaluate the relationship between activity reported on EMA (ie, within-person effect) and neurocognitive performance (ie, mCWIT and mVLT performance) as well as the percentage of surveys in which each activity was reported and the average neurocognitive performance (ie, the between-person effect). Two sets of analyses were completed: (1) same-survey associations, with activity selected on the EMA survey as a predictor of neurocognitive performance at the same survey (ie, mCWIT and mVLT performance), and (2) lagged associations, with activity at the previous survey within the same day as a predictor of next-survey neurocognitive performance. All models controlled for HIV status and study day (1-14, centered to day 1). Age, sex, education, and race or ethnicity (non-Hispanic White vs all other race or ethnicities) were initially included in the models if they were associated with the mobile cognitive test score at $P<.10$, and were retained as covariates if they remained associated with the outcome at $P<.10$. Therefore, mCWIT analyses included age and race or ethnicity as covariates, and mVLT analyses included education as a covariate. All analyses were reexamined with depressed mood as a covariate and excluding surveys in which participants reported alcohol or marijuana use. Results were considered statistically significant at $P<.05$. R software (version 3.6.1) was used for all analyses, and the lme4 package for R was used for mixed-effects regression analyses.

Table 2. Ecological momentary assessment and mobile cognitive test variables by HIV serostatus (N=103).

Variable	HIV+ (n=67)	HIV− (n=36)	*t* value or chi-square	*P* value
EMA[a] variables, mean (SD)				
Number of EMA surveys completed	91 (9.0)	90 (8.5)	0.3	.80
Percent of cognitively stimulating activities	16 (15.5)	19 (13.2)	−1.0	.09
Percent of passive leisure activities	35 (17.2)	29 (14.0)	1.7	.15
Percent of IADLs[b]	19 (8.1)	22 (9.0)	−1.7	.08
Mobile cognitive testing variables				
Average mCWIT[c] total seconds, mean (SD)	24.0 (6.7)	21.6 (4.2)	2.2	.03
Percent mCWIT completed, mean % (SD)	84 (18.5)	84 (20.0)	−0.1	.95
Average mVLT[d] total words, mean (SD)	19.7 (4.9)	21.1 (3.8)	−1.4	.15
Percent mVLT completed, mean % (SD)	79 (18.9)	82 (13.3)	−0.9	.39

[a]EMA: ecological momentary assessment.

[b]IADLs: instrumental activities of daily living.

[c]mCWIT: mobile Color-Word Interference Test.

[d]mVLT: mobile Verbal Learning Test.

Results

Sample Characteristics

Participants' demographic and clinical characteristics by HIV serostatus are presented in Table 1. On average, participants in this study were 59 years old (SD 6.4; range 50-74 years) with 14 years of education; 63% (65/103) of the participants were non-Hispanic White and only 34% (34/103) reported being employed (part or full time). Persons with HIV were more likely to be male (79% HIV+ vs 56% HIV−; $P=.01$) with slightly less education (13.9 years HIV+ vs 15.0 years HIV−; $P=.04$). Persons with HIV had high rates of current ART use 94% (63/67), and almost all persons with HIV had an undetectable viral load 97% (60/62).

On average, participants had high survey adherence, with an average of 91% (IQR 88%-96%) of EMA surveys completed, 85% of mCWIT trials (IQR 79%-100%), and 80% of mVLT trials (IQR 71%-93%). As EMA adherence ranged from 57% to 100%, no participants were excluded because of low adherence to the mobile EMA. On average, participants reported engaging in a cognitively stimulating activity on 17% of the surveys (range 0%-61%). Participants reported performing a passive leisure activity on 33% of surveys (range 4%-79%), which was predominantly watching television (19% of total surveys on average). IADLs were reported in 20% of the surveys (range 0%-41%). The percentage of activity types did not significantly differ by HIV serostatus. The HIV− group performed better on the mCWIT (HIV+: 24.0 seconds on average vs HIV−: 21.6 seconds on average; P=.03); however, there was no significant difference in average performance on the mVLT by HIV serostatus (HIV+: 19.7 words on average vs HIV−: 21.2 words on average; P=.15). EMA variables and mobile cognitive testing variables by HIV serostatus are shown in Table 2.

Same-Survey Associations: Activity as a Predictor of Same-Survey Neurocognitive Performance

Table 3 displays the linear mixed-effects models for all associations between activity reported on the EMA survey questions and same-survey mobile cognitive testing (ie, mCWIT and mVLT performance). Engagement in a cognitively stimulating activity (or *just prior* engagement if they stopped the activity to take the test) was significantly associated with taking less time to complete the mCWIT (β=−1.1; P=.007), whereas engagement in a passive leisure activity in the same survey as the mCWIT was significantly associated with taking longer to complete the mCWIT (β=.9; P=.005). Engaging in IADLs was not significantly associated with mCWIT performance. There was no between-person effect, as the percentage of all study surveys in which participants engaged in any of the activity types was not significantly associated with mCWIT performance (P>.05).

Table 3. Mixed-effects models for associations between activity and same-survey cognition.

Model	Estimate	95% CI	*P* value
Mobile Color-Word Interference Test[a]			
Model 1: Cognitively stimulating activities			
Cognitively stimulating activity (reference group: not doing activity)[b]	−1.117	−1.934 to −0.305	.007
Percent cognitively stimulating activities[c]	0.009	−0.066 to 0.084	.82
Model 2: Passive leisure activities			
Passive leisure activity (reference group: not doing activity)[b]	0.942	0.294 to 1.595	.005
Percent passive leisure activities[c]	0.032	−0.034 to 0.099	.35
Model 3: Instrumental activities of daily living			
Instrumental activities of daily living (reference group: not doing activity)[b]	−0.402	−1.170 to 0.369	.31
Percent instrumental activities of daily living[c]	−0.004	−0.134 to 0.127	.96
Mobile Verbal Learning Test[d]			
Model 4: Cognitively stimulating activities			
Cognitively stimulating activity (reference group: not doing activity)[b]	0.313	−0.334 to 0.961	.34
Percent cognitively stimulating activities[c]	0.070	0.011 to 0.129	.02
Model 5: Passive leisure activities			
Passive leisure activity (reference group: not doing activity)[b]	0.418	−0.103 to 0.937	.12
Percent passive leisure activities[c]	-0.070	−0.123 to −0.016	.01
Model 6: Instrumental activities of daily living			
Instrumental activities of daily living (reference group: not doing activity)[b]	−0.255	−0.820 to 0.310	.38
Percent instrumental activities of daily living[c]	0.009	−0.090 to 0.110	.86

[a]Mobile Color-Word Interference Test analyses controlled for study day, HIV status, age, and race or ethnicity.

[b]Modeled as a within-person variable.

[c]Modeled as between-person variables.

[d]Mobile Verbal Learning Test analyses controlled for study day, HIV status, and education.

The same-survey activity was not significantly related to mVLT performance within persons. Between persons, however, a higher percentage of surveys in which individuals reported having been engaged in cognitively stimulating activities was significantly associated with recalling more words on the mVLT on average (β=.07; *P*=.02). Upon further examination, participants in the top quartile of cognitively stimulating activities performed significantly better than participants in the bottom quartile (21.6 average mVLT words in top quartile of engagement in cognitively stimulating activities vs 17.8 average mVLT words in the bottom quartile of cognitively stimulating activities; *P*=.003; Cohen *d*=0.91). In contrast, reporting more passive leisure activities was significantly associated with fewer words recalled on the mVLT on average (β=−.07; *P*=.01). Similar to the findings with cognitively stimulating activities, participants in the lowest quartile for passive leisure activities performed significantly better on the mVLT than those in the top quartile (21.6 average mVLT words in the bottom quartile of engagement in passive leisure activities vs 17.9 average mVLT words in the top quartile of engagement in passive leisure activities; *P*=.004; Cohen *d*=0.87). The percentage of IADL activities was not significantly associated with mVLT performance. In follow-up analyses, including depressed mood as a time-varying covariate or excluding instances of alcohol and marijuana use in these models did not significantly impact any of the mCWIT or mVLT associations.

Lagged Associations: Activity as a Predictor of Next-Survey Cognition

Table 4 shows all linear mixed-effects models in which activity (ie, doing activity or not at a given survey) predicts cognitive performance on the next EMA survey within the same day (ie, survey 1 activity predicting survey 2 mobile cognitive test, survey 2 activity predicting survey 3 mobile cognitive test, and survey 3 activity predicting survey 4 mobile cognitive test). There were no significant associations between activity and the mCWIT or mVLT at the next EMA survey. Including depressed mood as a time-varying covariate or excluding instances of alcohol and marijuana use did not change these results.

Table 4. Mixed-effects models for associations between activity and next-survey cognition.

Model	Estimate	95% CI	*P* value
Mobile Color-Word Interference Test[a]			
Model 1: Cognitively stimulating activities			
Cognitively stimulating activity (reference group: not doing activity)[b]	0.510	−0.878 to 0.547	.39
Model 2: Passive leisure activities			
Passive leisure activity (reference group: not doing activity)[b]	0.449	−0.261 to 1.025	.38
Model 3: Instrumental activities of daily living			
Instrumental activities of daily living (reference group: not doing activity)[b]	−0.762	−0.230 to 1.083	.13
Mobile Verbal Learning Test[c]			
Model 4: Cognitively stimulating activities			
Cognitively stimulating activity (reference group: not doing activity)[b]	−0.167	−0.634 to 1.666	.65
Model 5: Passive leisure activities			
Passive leisure activity (reference group: not doing activity)[b]	0.385	−0.545 to 1.440	.24
Model 6: Instrumental activities of daily living			
Instrumental activities of daily living (reference group: not doing activity)[b]	−0.255	−1.757 to 0.236	.38

[a]Mobile Color-Word Interference Test analyses controlled for study day, HIV status, age, and race or ethnicity.

[b]Modeled as a within-person variable.

[c]Mobile Verbal Learning Test analyses controlled for study day, HIV status, and education.

Discussion

Principal Findings

This is one of the first studies to examine real-time relationships between activity reported on an EMA survey and mobile cognitive test performance in middle-aged and older adults in naturalistic environments. A total of 91% of EMA surveys were completed, and approximately 80% of mobile cognitive tests had valid data, thus demonstrating high adherence to the study protocol. In addition, this study demonstrates the utility of mobile cognitive testing for frequent monitoring of neurocognitive abilities to study the dynamic relationships between cognition and aspects of daily life.

Overall, we found that engagement in cognitively stimulating activities was associated with better executive functioning (mCWIT) and verbal learning (mVLT), whereas engagement in passive leisure activities was associated with worse executive functioning and verbal learning. In addition, these relationships did not seem to be explained by depressed mood or substance use (ie, alcohol or marijuana). Interestingly, it was the

same-survey relationship (ie, reporting engagement in a passive leisure or cognitively stimulating activity on the same survey as taking the mCWIT) that was associated with executive functioning within persons. In both tests, the observed effects were small. For example, engagement in cognitively stimulating or passive leisure activities on average related to only about a one-second within-person difference on the mCWIT. Similarly, on the mVLT, there was only a one-word difference for a 15% difference in the percentage of cognitively stimulating activities or a 15% difference in passive leisure activities. It is possible that in daily life this effect is negligible; however, these minor differences may be more apparent in *real-world* tasks that require longer, sustained attention. Although few studies have examined the relationship between activity and cognition, the limited research does suggest that some activities such as cognitively stimulating activities, socializing, and physical activity can provide *cognitive boosts*; however, similar to this study, effect sizes are usually small [24,36-38]. Therefore, this study adds to these emerging findings by (1) suggesting that cognitively stimulating activities are associated with better neurocognitive performance and (2) being one of the first to suggest that passive leisure activities may be associated with worse executive functioning performance in middle-aged and older adults with and without HIV in the same time frame. Therefore, this observational study supports that there may be an association between activity and cognitive function, thus suggesting that real-time interventions should be investigated to examine if these interventions may yield clinically meaningful results.

An important point to consider is that, by design, the mobile cognitive tests were taken in nonstandardized environments, and performance may be impacted by other factors in the environment, such as ambient noise or multitasking. Although it is these very factors that may make mobile cognitive testing more ecologically valid, it is possible that specific activities may be related to a greater chance of distractions in the environment (eg, watching television and not turning it off while taking the test). Therefore, we cannot confirm the mechanisms underlying better or worse performance on the mobile cognitive tests (eg, distraction vs a neurobehavioral process affecting cognition).

Conversely, activities did not appear to have a real-time association with verbal learning within persons; rather, the best predictor of mVLT performance was simply the total percentage of activities an individual reported over the entire study period. The relationship between more sedentary or passive activities such as watching television and worse neurocognitive performance has been observed in both aging [14,39,40] and HIV studies [18,41] and may be because of a number of factors. For example, high levels of sedentary behavior have been linked to worse cardiovascular health, which is associated with worse neurocognitive functioning [42,43]. In addition, increased television time is also associated with psychological factors such as depression and social isolation, which have also been associated with worse neurocognitive functioning [44,45]. We speculate that the relationship between passive activities and worse learning could also be bidirectional, as those with worse overall neurocognitive functioning may be more likely to engage in more passive activities. We also found that reporting more cognitively stimulating activities was associated with better verbal learning on average. The relationship between cognitively engaging activities and better cognition has also been documented in both older adults and persons with HIV and is thought to be related to increased cognitive reserve that can be protective against neurocognitive aging and HIV-associated neurocognitive impairment [41,46,47].

There are many possible reasons that could contribute to the observed difference, whereby the executive functioning task was associated with the current activity within persons, whereas the learning task was associated with the overall percentage of cognitively stimulating and passive leisure activities. One possible explanation may be that the mCWIT was a timed test, whereas the mVLT was not. Therefore, we could speculate that processing speed may have been more affected by one's surrounding environment and/or activity rather than executive functioning more specifically. In addition, it may be that the association between cognitively stimulating or passive leisure activities and executive functioning is more transient, whereas it is the accumulative effect of the different activities that is associated with verbal learning.

Finally, for both mobile cognitive tests, we did not observe any significant lagged effects such that activity on the previous survey (approximately 3-4 hours before) was not significantly related to cognition at the next time point. This suggests that these relationships may not be long lasting and/or that other activities in the interim may wash out the effect. In addition, because we chose to restrict the lagged analyses within the same day and therefore did not examine the relationship between activity the night prior and mobile cognitive testing in the morning, the lagged analyses did not examine morning cognition and thus may not reflect cognition throughout the entire day.

Limitations

There are additional limitations to this study that should be considered when interpreting the results as well as to improve future research. First, it is possible that within each activity type, some activities may be more beneficial than others. For example, within IADLs, working on finances may be more cognitively stimulating than riding in a car or taking public transportation. Due to limited occurrences of specific activities, we were unable to examine the association with more specific activities. Future studies with larger sample sizes or longer monitoring periods are needed to address this limitation in the literature. Second, we do not have additional information on each specific activity in which there is likely variability; for example, we do not know what type of television programs were watched. Moreover, participants were only able to select 1 activity but may have been engaged in multiple activities (eg, watching television and cooking a meal); thus, forcing participants to choose the primary activity in which they were engaged. Future research may want to allow participants to select multiple activities and examine the impact of multitasking on functioning. Third, while the majority of participants had excellent adherence to the EMA surveys, it is possible that the proportions of activities are somewhat biased. For example, certain activities that require more cognitive demand or attention

(eg, physical activity and socializing) may be associated with a greater likelihood of missing a survey, and therefore bias proportion of reported activities. Fourth, because of the low rate at which social activities and physical activity were endorsed, we were unable to examine these activities. Future research aimed at understanding the cognitive impact of physical and social activity may benefit from more frequent surveys, querying about all activities since the last survey (eg, as done by Granholm et al [48]), or integrating passive assessments of other behaviors (eg, continuous monitoring of physical activity via actigraphy). Finally, this study included a large percentage of persons with HIV who were relatively healthy, with high rates of ART use and viral suppression. Therefore, these findings may not be generalizable to all middle-aged and older adults or to all persons with HIV.

Conclusions

This study demonstrates that it is possible to assess momentary fluctuations in cognition in relation to real-time activities in participants' lived environments. This research methodology is particularly advantageous, given that it can be completed in participants' own environments and does not require face-to-face contact. As a timely example, these ambulatory assessment methods could be used to investigate and track the neurocognitive changes in people recovering from COVID-19 [49].

Overall, the observed effects in this study were small, but they did suggest that cognitively stimulating activities just before testing were associated with better performance on mobile cognitive tests, whereas passive leisure activities were associated with worse performance. These results demonstrate that more research is needed to understand the contexts (such as environment, biological processes, or both) that drive these relationships to develop better recommendations and interventions to boost neurocognitive functioning. Digital health technologies may be particularly useful intervention tools, and these interventions may be particularly beneficial for older adults and older persons with HIV at greater risk for neurocognitive deficits than the general public. Given that this study only examined verbal learning and executive functioning, additional research should examine other neurocognitive domains and examine which neurocognitive domains may be most responsive to short-term variation in activities versus which are more responsive to the accumulative effects of different activities to inform future interventional research.

Acknowledgments

This work was supported by the National Institutes of Health: RM, grant numbers NIMH K23 MH105297, NIMH K23 MH107260 S1, NIMH R21 MH116104, NIA AG062387; LC, grant number NIDA T32 DA031098; and EP, grant number NIAAA F31 AA027198. The HIV Neurobehavioral Research Center is supported by Center award P30MH062512 from NIMH. The San Diego HIV Neurobehavioral Research Center (HNRC) group is affiliated with the University of California, San Diego, the Naval Hospital, San Diego, and the Veterans Affairs San Diego Healthcare System, and includes: Director: Robert K Heaton, PhD, Co-Director: Igor Grant, MD; Associate Directors: J Hampton Atkinson, MD, Ronald J Ellis, MD, PhD, and Scott Letendre, MD; Center Manager: Jennifer Iudicello, PhD; Donald Franklin, Jr; Melanie Sherman; NeuroAssessment Core: Ronald J Ellis, MD, PhD (PI), Scott Letendre, MD, Thomas D Marcotte, PhD, Christine Fennema-Notestine, PhD, Debra Rosario, MPH, Matthew Dawson; NeuroBiology Core: Cristian Achim, MD, PhD (PI), Ana Sanchez, PhD, Adam Fields, PhD; NeuroGerm Core: Sara Gianella Weibel, MD (PI), David M Smith, MD, Rob Knight, PhD, Scott Peterson, PhD; Developmental Core: Scott Letendre, MD (PI), J Allen McCutchan; Participant Accrual and Retention Unit: J Hampton Atkinson, MD (PI) Susan Little, MD, Jennifer Marquie-Beck, MPH; Data Management and Information Systems Unit: Lucila Ohno-Machado, PhD (PI), Clint Cushman; Statistics Unit: Ian Abramson, PhD (PI), Florin Vaida, PhD (Co-PI), Anya Umlauf, MS, Bin Tang, MS. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the United States Government.

Conflicts of Interest

RM is a cofounder of KeyWise AI, Inc and a consultant for NeuroUX. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Neuropsychological battery by domain.
[DOCX File , 14 KB - mhealth_v8i9e19579_app1.docx]

References

1. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). World Health Organization. 2020. URL: https://tinyurl.com/rjdtx2k [accessed 2020-04-01]
2. Marcotte T, Grant I, editors. Neuropsychology of Everyday Functioning. New York, USA: Guilford Press; 2009.
3. Marcotte T, Scott JC, Kamat R, Heaton RK. Neuropsychology and the prediction of everyday functioning. In: Marcotte T, Grant I, editors. Neuropsychology of Everyday Functioning. New York, USA: Guilford Press; 2009:5-38.
4. Howieson D. Current limitations of neuropsychological tests and assessment procedures. Clin Neuropsychol 2019 Feb;33(2):200-208. [doi: 10.1080/13854046.2018.1552762] [Medline: 30608020]

5. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 5;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
6. Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
7. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging 2019 Jan;3(1):igy038 [FREE Full text] [doi: 10.1093/geroni/igy038] [Medline: 30619948]
8. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int J Methods Psychiatr Res 2017 Dec;26(4):e1562 [FREE Full text] [doi: 10.1002/mpr.1562] [Medline: 28370881]
9. Germine L, Reinecke K, Chaytor NS. Digital neuropsychology: challenges and opportunities at the intersection of science and software. Clin Neuropsychol 2019 Feb;33(2):271-286. [doi: 10.1080/13854046.2018.1535662] [Medline: 30614374]
10. Moore RC, Campbell LM, Delgadillo JD, Paolillo EW, Sundermann EE, Holden J, et al. Smartphone-based measurement of executive function in older adults with and without HIV. Arch Clin Neuropsychol 2020 May 22;35(4):347-357. [doi: 10.1093/arclin/acz084] [Medline: 31942632]
11. Bouvard A, Dupuy M, Schweitzer P, Revranche M, Fatseas M, Serre F, et al. Feasibility and validity of mobile cognitive testing in patients with substance use disorders and healthy controls. Am J Addict 2018 Oct;27(7):553-556. [doi: 10.1111/ajad.12804] [Medline: 30260085]
12. Schweitzer P, Husky M, Allard M, Amieva H, Pérès K, Foubert-Samier A, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. Int J Methods Psychiatr Res 2017 Sep;26(3):e1521 [FREE Full text] [doi: 10.1002/mpr.1521] [Medline: 27539327]
13. Wilson RS, Segawa E, Boyle PA, Bennett DA. Influence of late-life cognitive activity on cognitive health. Neurology 2012 Apr 10;78(15):1123-1129 [FREE Full text] [doi: 10.1212/WNL.0b013e31824f8c03] [Medline: 22491864]
14. Bakrania K, Edwardson C, Khunti K, Bandelow S, Davies M, Yates T. Associations between sedentary behaviors and cognitive function: cross-sectional and prospective findings from the UK biobank. Am J Epidemiol 2018 Mar 1;187(3):441-454. [doi: 10.1093/aje/kwx273] [Medline: 28992036]
15. Hoang TD, Reis J, Zhu N, Jacobs DR, Launer LJ, Whitmer RA, et al. Effect of early adult patterns of physical activity and television viewing on midlife cognitive function. JAMA Psychiatry 2016 Jan;73(1):73-79 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.2468] [Medline: 26629780]
16. Blasko I, Jungwirth S, Kemmler G, Weissgram S, Tragl K, Fischer P. Leisure time activities and cognitive functioning in middle European population-based study. Eur Geriatr Med 2014 Jun;5(3):200-207. [doi: 10.1016/j.eurger.2013.09.003]
17. Fancourt D, Steptoe A. Television viewing and cognitive decline in older age: findings from the English longitudinal study of ageing. Sci Rep 2019 Feb 28;9(1):2851 [FREE Full text] [doi: 10.1038/s41598-019-39354-4] [Medline: 30820029]
18. Moore RC, Kaufmann CN, Rooney AS, Moore DJ, Eyler LT, Granholm E, et al. Feasibility and acceptability of ecological momentary assessment of daily functioning among older adults with HIV. Am J Geriatr Psychiatry 2017 Aug;25(8):829-840 [FREE Full text] [doi: 10.1016/j.jagp.2016.11.019] [Medline: 28065496]
19. Verghese J, Lipton RB, Katz MJ, Hall CB, Derby CA, Kuslansky G, et al. Leisure activities and the risk of dementia in the elderly. N Engl J Med 2003 Jun 19;348(25):2508-2516. [doi: 10.1056/NEJMoa022252] [Medline: 12815136]
20. Fazeli PL, Woods SP, Heaton RK, Umlauf A, Gouaux B, Rosario D, HNRP Group. An active lifestyle is associated with better neurocognitive functioning in adults living with HIV infection. J Neurovirol 2014 Jun;20(3):233-242 [FREE Full text] [doi: 10.1007/s13365-014-0240-z] [Medline: 24554483]
21. Farias ST, Mungas D, Jagust W. Degree of discrepancy between self and other-reported everyday functioning by cognitive status: dementia, mild cognitive impairment, and healthy elders. Int J Geriatr Psychiatry 2005 Sep;20(9):827-834 [FREE Full text] [doi: 10.1002/gps.1367] [Medline: 16116577]
22. Thames AD, Becker BW, Marcotte TD, Hines LJ, Foley JM, Ramezani A, et al. Depression, cognition, and self-appraisal of functional abilities in HIV: an examination of subjective appraisal versus objective performance. Clin Neuropsychol 2011 Feb;25(2):224-243 [FREE Full text] [doi: 10.1080/13854046.2010.539577] [Medline: 21331979]
23. Harvey PD, Velligan DI, Bellack AS. Performance-based measures of functional skills: usefulness in clinical treatment studies. Schizophr Bull 2007 Sep;33(5):1138-1148 [FREE Full text] [doi: 10.1093/schbul/sbm040] [Medline: 17493956]
24. Allard M, Husky M, Catheline G, Pelletier A, Dilharreguy B, Amieva H, et al. Mobile technologies in the early detection of cognitive decline. PLoS One 2014;9(12):e112197 [FREE Full text] [doi: 10.1371/journal.pone.0112197] [Medline: 25536290]
25. Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of cognitive impairment without dementia in the United States. Ann Intern Med 2008 Mar 18;148(6):427-434 [FREE Full text] [doi: 10.7326/0003-4819-148-6-200803180-00005] [Medline: 18347351]
26. Makinson A, Dubois J, Eymard-Duvernay S, Leclercq P, Zaegel-Faucher O, Bernard L, et al. Increased prevalence of neurocognitive impairment in aging people living with human immunodeficiency virus: the ANRS EP58 hand 55-70 study. Clin Infect Dis 2020 Jun 10;70(12):2641-2648. [doi: 10.1093/cid/ciz670] [Medline: 31755936]

27. Heaton RK, Clifford DB, Franklin DR, Woods SP, Ake C, Vaida F, CHARTER Group. HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER study. Neurology 2010 Dec 7;75(23):2087-2096 [FREE Full text] [doi: 10.1212/WNL.0b013e318200d727] [Medline: 21135382]
28. Jeste DV, Palmer BW, Appelbaum PS, Golshan S, Glorioso D, Dunn LB, et al. A new brief instrument for assessing decisional capacity for clinical research. Arch Gen Psychiatry 2007 Aug;64(8):966-974. [doi: 10.1001/archpsyc.64.8.966] [Medline: 17679641]
29. World Health Organization. Composite International Diagnostic Interview (CIDI, version 2.1). Geneva, Switzerland: World Health Organization; 1997.
30. Cysique LA, Franklin D, Abramson I, Ellis RJ, Letendre S, Collier A, CHARTER Group, HNRC Group. Normative data and validation of a regression based summary score for assessing meaningful neuropsychological change. J Clin Exp Neuropsychol 2011 Jun;33(5):505-522 [FREE Full text] [doi: 10.1080/13803395.2010.535504] [Medline: 21391011]
31. Carey CL, Woods SP, Gonzalez R, Conover E, Marcotte TD, Grant I, HNRC Group. Predictive validity of global deficit scores in detecting neuropsychological impairment in HIV infection. J Clin Exp Neuropsychol 2004 May;26(3):307-319. [doi: 10.1080/13803390490510031] [Medline: 15512922]
32. Wilkinson GS, Robertson GJ. Wide Range Achievement Test (WRAT4). Lutz, FL: Psychological Assessment Resources; 2006.
33. Hinkin CH, Castellon SA, Hardy DJ, Granholm E, Siegle G. Computerized and traditional stroop task dysfunction in HIV-1 infection. Neuropsychology 1999 Apr;13(2):306-316. [doi: 10.1037//0894-4105.13.2.306] [Medline: 10353380]
34. Iudicello JE, Woods SP, Deutsch R, Grant I, HIV Neurobehavioral Research Program Hnrp Group. Combined effects of aging and HIV infection on semantic verbal fluency: a view of the cortical hypothesis through the lens of clustering and switching. J Clin Exp Neuropsychol 2012;34(5):476-488 [FREE Full text] [doi: 10.1080/13803395.2011.651103] [Medline: 22292479]
35. Benedict RH, Schretlen D, Groninger L, Brandt J. Hopkins verbal learning test – revised: normative data and analysis of inter-form and test-retest reliability. Clin Neuropsychol 1998;12(1):43-55. [doi: 10.1076/clin.12.1.43.1726]
36. Ybarra O, Winkielman P, Yeh I, Burnstein E, Kavanagh L. Friends (and sometimes enemies) with cognitive benefits. Soc Psychol Pers Sci 2010 Oct 13;2(3):253-261. [doi: 10.1177/1948550610386808]
37. Ybarra O, Burnstein E, Winkielman P, Keller MC, Manis M, Chan E, et al. Mental exercising through simple socializing: social interaction promotes general cognitive functioning. Pers Soc Psychol Bull 2008 Feb;34(2):248-259. [doi: 10.1177/0146167207310454] [Medline: 18212333]
38. Chang YK, Labban JD, Gapin JI, Etnier JL. The effects of acute exercise on cognitive performance: a meta-analysis. Brain Res 2012 May 9;1453:87-101. [doi: 10.1016/j.brainres.2012.02.068] [Medline: 22480735]
39. Kesse-Guyot E, Charreire H, Andreeva VA, Touvier M, Hercberg S, Galan P, et al. Cross-sectional and longitudinal associations of different sedentary behaviors with cognitive performance in older adults. PLoS One 2012;7(10):e47831 [FREE Full text] [doi: 10.1371/journal.pone.0047831] [Medline: 23082222]
40. Hamer M, Stamatakis E. Prospective study of sedentary behavior, risk of depression, and cognitive impairment. Med Sci Sports Exerc 2014 Apr;46(4):718-723 [FREE Full text] [doi: 10.1249/MSS.0000000000000156] [Medline: 24121248]
41. Milanini B, Ciccarelli N, Fabbiani M, Limiti S, Grima P, Rossetti B, et al. Cognitive reserve and neuropsychological functioning in older HIV-infected people. J Neurovirol 2016 Oct;22(5):575-583. [doi: 10.1007/s13365-016-0426-7] [Medline: 26965299]
42. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia 2012 Nov;55(11):2895-2905. [doi: 10.1007/s00125-012-2677-z] [Medline: 22890825]
43. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
44. Boss L, Kang D, Branson S. Loneliness and cognitive function in the older adult: a systematic review. Int Psychogeriatr 2015 Apr;27(4):541-553. [doi: 10.1017/S1041610214002749] [Medline: 25554219]
45. Teychenne M, Ball K, Salmon J. Sedentary behavior and depression among adults: a review. Int J Behav Med 2010 Dec;17(4):246-254. [doi: 10.1007/s12529-010-9075-z] [Medline: 20174982]
46. Yates LA, Ziser S, Spector A, Orrell M. Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis. Int Psychogeriatr 2016 Nov;28(11):1791-1806. [doi: 10.1017/S1041610216001137] [Medline: 27502691]
47. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol 2012 Nov;11(11):1006-1012 [FREE Full text] [doi: 10.1016/S1474-4422(12)70191-6] [Medline: 23079557]
48. Granholm E, Holden JL, Mikhael T, Link PC, Swendsen J, Depp C, et al. What do people with schizophrenia do all day? Ecological momentary assessment of real-world functioning in schizophrenia. Schizophr Bull 2020 Feb 26;46(2):242-251. [doi: 10.1093/schbul/sbz070] [Medline: 31504955]

49. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]

Abbreviations

ADL: activities of daily living
ART: antiretroviral therapy
EMA: ecological momentary assessment
EMCA: ecological momentary cognitive assessment
HNRP: HIV Neurobehavioral Research Program
IADL: instrumental activities of daily living
mCWIT: mobile Color-Word Interference Test
mVLT: mobile Verbal Learning Test
NIMH: National Institute of Mental Health
OS: operating system
UCSD: University of California, San Diego
WRAT-4: Wide Range Achievement Test, fourth edition Reading Subtest

Edited by G Eysenbach; submitted 23.04.20; peer-reviewed by A Wong, MJ Brouillette, A Dargel; comments to author 12.06.20; revised version received 05.08.20; accepted 06.08.20; published 24.09.20.

Please cite as:
Campbell LM, Paolillo EW, Heaton A, Tang B, Depp CA, Granholm E, Heaton RK, Swendsen J, Moore DJ, Moore RC
Daily Activities Related to Mobile Cognitive Performance in Middle-Aged and Older Adults: An Ecological Momentary Cognitive Assessment Study
JMIR Mhealth Uhealth 2020;8(9):e19579
URL: http://mhealth.jmir.org/2020/9/e19579/
doi:10.2196/19579
PMID:32969829

Original Paper

Mobile Health Apps on COVID-19 Launched in the Early Days of the Pandemic: Content Analysis and Review

Long Chiau Ming[1*], PhD; Noorazrina Untong[1*], BHScPharm; Nur Amalina Aliudin[1*], BHScPharm; Norliza Osili[1*], BHScPharm; Nurolaini Kifli[1], PhD; Ching Siang Tan[2], PhD; Khang Wen Goh[3], PhD; Pit Wei Ng[4], BPharm (Hons), BCPS, BCACP; Yaser Mohammed Al-Worafi[5,6], PhD; Kah Seng Lee[7], MSc; Hui Poh Goh[1], PhD

[1]Pengiran Anak Puteri Rashidah Sa'adatul Bolkiah Institute of Health Sciences, Universiti Brunei Darussalam, Gadong, Brunei Darussalam

[2]School of Pharmacy, KPJ Healthcare University College, Nilai, Malaysia

[3]Faculty of Science and Technology, Quest International University Perak, Ipoh, Malaysia

[4]Department of Pharmacy, National University Health System, Singapore, Singapore

[5]College of Pharmacy, University of Science and Technology of Fujairah, Fujairah, United Arab Emirates

[6]College of Pharmacy, University of Science and Technology, Sana'a, Yemen

[7]Faculty of Pharmacy, University of Cyberjaya, Cyberjaya, Malaysia

[*]these authors contributed equally

Corresponding Author:
Long Chiau Ming, PhD
Pengiran Anak Puteri Rashidah Sa'adatul Bolkiah Institute of Health Sciences
Universiti Brunei Darussalam
Tungku Link Rd
Gadong, BE1410
Brunei Darussalam
Phone: 673 2460922 ext 2202
Fax: 673 2461081
Email: longchiauming@gmail.com

Abstract

Background: Mobile health (mHealth) app use is a major concern because of the possible dissemination of misinformation that could harm the users. Particularly, it can be difficult for health care professionals to recommend a suitable app for coronavirus disease (COVID-19) education and self-monitoring purposes.

Objective: This study aims to analyze and evaluate the contents as well as features of COVID-19 mobile apps. The findings are instrumental in helping health care professionals to identify suitable mobile apps for COVID-19 self-monitoring and education. The results of the mobile apps' assessment could potentially help mobile app developers improve or modify their existing mobile app designs to achieve optimal outcomes.

Methods: The search for the mHealth apps available in the android-based Play Store and the iOS-based App Store was conducted between April 18 and May 5, 2020. The region of the App Store where we performed the search was the United States, and a virtual private network app was used to locate and access COVID-19 mobile apps from all countries on the Google Play Store. The inclusion criteria were apps that are related to COVID-19 with no restriction in language type. The basic features assessment criteria used for comparison were the requirement for free subscription, internet connection, education or advisory content, size of the app, ability to export data, and automated data entry. The functionality of the apps was assessed according to knowledge (information on COVID-19), tracing or mapping of COVID-19 cases, home monitoring surveillance, online consultation with a health authority, and official apps run by health authorities.

Results: Of the 223 COVID-19–related mobile apps, only 30 (19.9%) found in the App Store and 28 (44.4%) in the Play Store matched the inclusion criteria. In the basic features assessment, most App Store (10/30, 33.3%) and Play Store (10/28, 35.7%) apps scored 4 out of 7 points. Meanwhile, the outcome of the functionality assessment for most App Store apps (13/30, 43.3%) was a score of 3 compared to android-based apps (10/28, 35.7%), which scored 2 (out of the maximum 5 points). Evaluation of the basic functions showed that 75.0% (n=36) of the 48 included mobile apps do not require a subscription, 56.3% (n=27) provide symptom advice, and 41.7% (n=20) have educational content. In terms of the specific functions, more than half of the included mobile apps are official mobile apps maintained by a health authority for COVID-19 information provision. Around 37.5% (n=18)

and 31.3% (n=15) of the mobile apps have tracing or mapping and home monitoring surveillance functions, respectively, with only 17% (n=8) of the mobile apps equipped with an online consultation function.

Conclusions: Most iOS-based apps incorporate infographic mapping of COVID-19 cases, while most android-based apps incorporate home monitoring surveillance features instead of providing focused educational content on COVID-19. It is important to evaluate the contents and features of COVID-19 mobile apps to guide users in choosing a suitable mobile app based on their requirements.

(JMIR Mhealth Uhealth 2020;8(9):e19796) doi:10.2196/19796

KEYWORDS

coronavirus; mobile medical app; self-care; mHealth; health education; app; COVID-19; content analysis

Introduction

The coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has received global attention. At the time of writing (April 2020), the number of confirmed cases across the world continues to rise. The World Health Organization announced COVID-19 as a pandemic on March 11, 2020 [1]. Although the SARS-CoV-2 pandemic has had a significant impact in terms of lockdown and mortality, the public has become more eager to obtain information about the spread of the infection [2]. Their efforts in keeping themselves up-to-date with the latest information about COVID-19 could involve existing technologies such as watching the national bulletin on the television or listening to the news on the radio. However, a majority of people may not like the idea of waiting for a live broadcast at a fixed schedule. Reading digital news articles and scrolling through reliable official websites seem to be the main option for tech-savvy individuals [3].

This opens up a golden opportunity for web or mobile medical app developers to create a platform for the public to provide them with the information they are looking for. With the advancement of mobile software and technology, mobile apps have become an important element in our daily life [4]. The use of mobile technology and devices has been found to be successful in the health care setting [5-8]. The term “mobile Health” (mHealth) has been used to describe any health care practice that is supported by mobile devices [9]. For instance, an mHealth app may help health care professionals in treating clinical diseases and educating patients on self-monitoring of the disease as well as reinforcing treatment adherence [10,11]. The use of mHealth apps has made health care and health information easily accessible [12]. Furthermore, the use of mHealth apps at the user’s convenience also helps to reduce the frequency of unnecessary hospital visits by stable patients, thus reducing the mobility of patients who are immunocompromised to high-risk areas [13,14].

The implementation of strategic features in mHealth that can help in diagnosis or symptom reporting has great potential in the management of infections. Additionally, the integration of relevant epidemiological data and geographical information of transmittable disease prevalence in a region will allow the tracing of cases, which can be used as an effective tool to control the spread of infection [15]. It is more effective to deliver health-related information through mHealth apps, as information can be exchanged rapidly and updated dynamically [10]. Mobile apps can potentially prevent the occurrence of a particular disease, as exchanged texts through a mobile app can promote communication, storage of information, and message delivery that drives users to make healthy lifestyle changes [11,16].

Recently the US Food and Drug Administration issued guidance and policy for mHealth apps to ensure their safety and effectiveness [17]. Meanwhile, other challenges revolve around information-sharing and transparency of services offered that could compromise the privacy of the app’s user [18,19]. mHealth app use is also a major concern among health care professionals because of the possible dissemination of misinformation that could harm the users or readers, as some information and services provided are not aligned with medical guidelines [20].

This study aims to analyze and evaluate the contents as well as features of COVID-19 mobile apps. The findings are instrumental in helping health care professionals to identify suitable mobile apps for COVID-19 self-monitoring and education. The results of the mobile apps assessment can potentially help mobile app developers improve or modify their existing mobile app designs to achieve optimal outcomes.

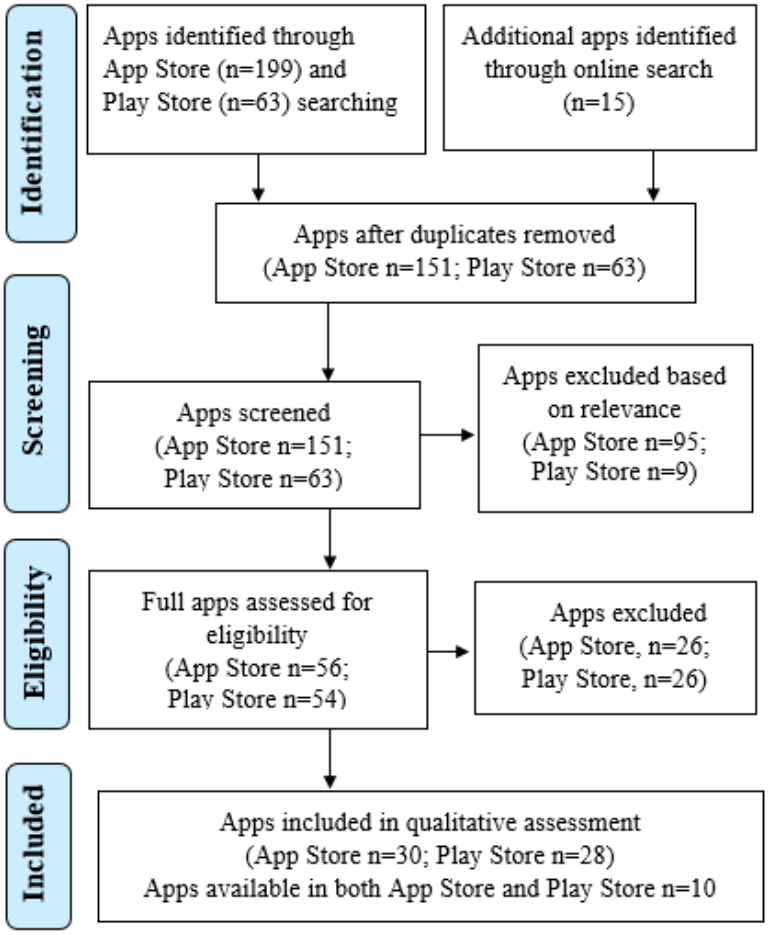
Methods

We performed a content analysis, comparison, and functionality assessment of selected mobile apps for COVID-19. First, a search for COVID-19 mobile apps was performed in two digital platforms: the App Store on the Apple iPhone 8 Plus and the Google Play Store on Oppo R9s and Vivo V9 smartphones. The search was conducted from April 18, 2020, to May 5, 2020. The region of the App Store where we performed the search was the United States, while a virtual private network (VPN) app named Touch VPN was used to locate and access COVID-19 mobile apps from all countries on the Google Play Store. The inclusion criteria to obtain relevant mHealth apps included apps launched for smartphone users and apps that are related to COVID-19 with no restriction in language type. The exclusion criteria include mobile apps that are launched on other devices such as iPads, tablets, and laptops; apps designed to provide quarantined users with their grocery or pharmacy supplies in response to containing the virus; and entrepreneurship apps designed to collect funds in support of organizations affected by COVID-19.

The keywords “Covid19,” “Coronavirus,” “Corona,” and “COVID-19” were used to find COVID-19 mobile apps in the App Store and the Play Store. To ensure that all relevant mobile apps were included, an online search on Google using the key

terms “mobile app,” “mHealth,” “Covid19,” “Coronavirus,” “Corona,” and “COVID-19” was also conducted. All mobile apps were then filtered according to the COVID-19 relevance and were further filtered according to the inclusion and exclusion criteria. The authors are mainly proficient in the English language, so only apps that support an English language user interface were assessed and reviewed. The summaries of the processes involved in selecting the relevant mobile apps from the App Store and Play Store are illustrated in Figure 1.

Figure 1. Selection process of mobile health apps in Apple's App Store and the Google Play Store.



The included mobile apps were assessed based on their basic features and functionalities. The basic features were modified from the outline of developed classification of mHealth apps evaluation criteria proposed by Nouri et al [21] and in the literature [11,16,22]. The included seven basic features were (1) no internet requirement, (2) size of app less than 50 MB, (3) no subscription needed (ie, free), (4) educational content (COVID-19 teaching), (5) export data (sharing of user’s data with other platforms), (6) automated data entry (automatic update of data without user interference), and (7) advisory function. Once the assessment of basic features was completed, the researchers convened again to categorize the apps into different groups according to their purpose and functionality, by reading the summary and explanation given by the developers of each included app. The categorized five functionalities of mobile apps were (1) knowledge (information on COVID-19), (2) tracing or mapping of COVID-19 cases, (3) home monitoring surveillance, (4) online consultation with a health authority, and (5) official mobile apps run by a health authority.

The basic features of all included mobile apps were screened individually by three researchers. Any disagreement was discussed until consensus was achieved. The full content of the included mobile apps were then individually examined by the same researchers. Any ambiguity was resolved by two senior researchers to confirm the functionality classification of all included mobile apps. One point was assigned to items that were fully satisfied. No point was given for each item that was partly satisfied or did not apply. There was a maximum of 7 and 5 points for the basic features and functionalities, respectively. Descriptive statistics (frequencies) were used to describe the characteristics of the apps according to the basic features and functionalities.

Results

The keywords used to search for mobile apps related to COVID-19 in Apple’s App Store and the Google Play Store gave a total of 151 and 63 apps, respectively. The apps were filtered according to the inclusion and exclusion criteria, giving a total of 30 mobile apps from the App Store and 28 mobile apps from the Play Store. The available mobile apps were grouped according to universal COVID-19 apps, country-specific apps, and non-English language apps, as

illustrated in Multimedia Appendix 1 for the Apple App Store and Multimedia Appendix 2 for the Google Play Store.

Out of the total number of mobile apps available with relevance to COVID-19, only 30 out of 151 (19.9%) that were found in the Apple App Store and 28 out of 63 (44.4%) in the Google Play Store were assessed. Selected mobile apps were assessed according to their basic features (Tables 1-3) and functionality (Tables 4-6). The results of the assessment follow a scoring system whereby the app was given a score of one for every criterion that it satisfied.

Table 1. Basic features assessment of mobile medical apps (iOS and android-based).

No	Name of mobile apps	No internet requirement	Size of app <50 MB	No subscription requirement (ie, free)	Educational content	Export data	Automated data entry	Advisory	Total score
Universal COVID-19[a] apps									
1	COVID Symptom Tracker	N/A[b]	✓	✓	N/A	✓	✓	✓	5
Country-specific apps									
2	BC COVID-19	N/A	✓	✓	✓	N/A	✓	✓	5
3	Canada COVID-19	N/A	✓	✓	✓	N/A	✓	✓	5
4	Coronavirus Australia	N/A	✓	✓	✓	✓	N/A	✓	5
5	COVA[c] Punjab	N/A	✓	N/A	—[d]	✓	✓	✓	4
6	HSE[e] COVID-19	N/A	✓	N/A	N/A	✓	✓	N/A	3
7	NCOVI	—	✓	✓	✓	—	—	✓	4
8	TraceTogether	N/A	✓	✓	N/A	✓	✓	✓	5
9	자가격리자 안전보호 (Self-Isolator Safety & Protection)	—	✓	✓	✓	—	—	✓	4
10	자가격리자 전담공무원 (Self-isolating Government Officials)	—	✓	✓	✓	—	—	✓	4

[a]COVID-19: coronavirus disease.

[b]N/A: not applicable.

[c]COVA: Corona Virus Alert.

[d]Absence of information regarding the feature in the app.

[e]HSE: Health Service Executive.

Table 2. Basic features assessment of mobile medical apps (iOS-based only).

No	Name of mobile apps	No internet requirement	Size of app <50 MB	No subscription requirement (ie, free)	Educational content	Export data	Automated data entry	Advisory	Total score
Universal COVID-19[a] apps									
1	APPLE COVID-19	✓	✓	✓	✓	N/A[b]	N/A	N/A	4
2	CDC[c]	N/A	✓	✓	✓	N/A	N/A	N/A	3
3	CoronaFACTS	N/A	✓	✓	✓	N/A	✓	N/A	4
4	Corona Checker	N/A	✓	✓	N/A	N/A	N/A	✓	3
5	COVID-19!	N/A	N/A	✓	✓	N/A	N/A	✓	3
6	HEALTHLYNKED COVID-19 TRACKER	N/A	✓	✓	✓	N/A	N/A	✓	4
7	RELIEF CENTRAL	N/A	✓	✓	✓	N/A	✓	✓	5
8	Patient Sphere COVID-19	N/A	✓	✓	N/A	✓	N/A	N/A	3
9	PreMedicus	—[d]	N/A	N/A	N/A	✓	N/A	✓	2
10	Mobile Angel Cancer Telemed	N/A	N/A	✓	N/A	✓	N/A	N/A	2
Country-specific apps									
11	BMC Combat Covid19	N/A	✓	✓	N/A	N/A	N/A	N/A	2
12	Corona-Care	N/A	N/A	N/A	N/A	✓	✓	✓	3
13	COVID-19 Gov PK[e]	N/A	✓	✓	✓	N/A	✓	✓	5
14	Covidom Patient	N/A	N/A	N/A	N/A	✓	N/A	✓	2
15	COVI QATAR	N/A	✓	✓	✓	N/A	N/A	✓	4
16	COVID-19 UAE[f]	N/A	N/A	✓	✓	N/A	✓	✓	4
17	CUREiTT	N/A	N/A	N/A	N/A	✓	✓	N/A	2
18	NJ COVID 19	N/A	✓	✓	✓	N/A	N/A	✓	4
19	STOP COVID19 CAT	N/A	✓	N/A	—	✓	—	—	2
20	Tarussud	N/A	✓	✓	✓	N/A	✓	✓	5

[a]COVID-19: coronavirus disease.

[b]N/A: not applicable.

[c]CDC: Centers for Disease Control and Prevention.

[d]Absence of information regarding the feature in the app.

[e]PK: Pakistan.

[f]UAE: United Arab Emirates.

Table 3. Basic features assessment of mobile medical apps (android-based only)

No	Name of mobile apps	No internet requirement	Size of app <50 MB	No subscription requirement (ie, free)	Educational content	Exported data	Automated data entry	Advisory	Total score
Universal COVID-19[a] apps									
1	Test Yourself Goa	N/A[b]	✓	✓	N/A	✓	N/A	✓	4
Country-specific apps									
2	Aarogya Setu	N/A	✓	✓	N/A	✓	✓	✓	5
3	CoBuddy – Covid19 Tool	N/A	✓	N/A	N/A	✓	✓	N/A	3
4	Corona Watch	N/A	✓	✓	N/A	N/A	✓	✓	4
5	COVI	N/A	✓	✓	✓	N/A	✓	✓	5
6	Covid-19	N/A	✓	N/A	N/A	✓	✓	N/A	3
7	COVID-19 NI[c]	✓	✓	✓	✓	N/A	✓	✓	6
8	COVID19 Feedback	N/A	✓	✓	N/A	✓	✓	N/A	4
9	COVID-19 Quarantine Monitor Tamil Nadu (official)	N/A	✓	✓	N/A	✓	✓	N/A	4
10	COVID-19 West Bengal Government	N/A	✓	✓	N/A	✓	✓	N/A	4
11	GCC[d] -Corona Monitoring	N/A	✓	✓	N/A	N/A	✓	N/A	3
12	GoK – Direct Kerala	N/A	✓	✓	N/A	N/A	N/A	N/A	2
13	Home Quarantine (Kwarantanna domowa)	N/A	N/A	✓	N/A	✓	✓	N/A	3
14	Mahakavach	N/A	✓	N/A	N/A	✓	✓	—[e]	3
15	MP[f] COVID RESPONSE APP	N/A	✓	✓	—	✓	—	—	3
16	Quarantine Watch	N/A	✓	N/A	—	✓	—	—	2
17	StayHomeSafe	✓	✓	N/A	—	✓	✓	—	4
18	Test Yourself Puducherry	N/A	✓	✓	✓	✓	✓	✓	6

[a]COVID-19: coronavirus disease.

[b]N/A: not applicable.

[c]NI: North Ireland.

[d]GCC: Greater Chennai Corporation.

[e]Absence of information regarding the feature in the app.

[f]MP: National Health Mission.

Table 4. Functionality assessment of mobile medical apps (iOS and android-based).

No	Name of mobile apps	Knowledge	Tracing/mapping of COVID-19[a] cases	Home monitoring surveillance	Online consultation with health authority	Official mobile app maintained by health authority	Total score
Universal COVID-19 apps							
1	COVID Symptom Tracker	✓	✓	N/A[b]	N/A	✓	3
Country-specific apps							
2	BC COVID-19	✓	N/A	N/A	N/A	✓	2
3	Canada COVID-19	✓	✓	N/A	N/A	✓	3
4	Coronavirus Australia	✓	✓	N/A	N/A	✓	3
5	COVA[c] Punjab	✓	✓	—[d]	✓	✓	4
6	HSE[e] COVID-19	✓	N/A	✓	N/A	✓	3
7	NCOVI	✓	✓	—	—	✓	3
8	TraceTogether	✓	✓	N/A	✓	✓	4
9	자가격리자 안전보호 (Self-Isolator Safety & Protection)	✓	—	✓	—	✓	3
10	자가격리자 전담공무원 (Self-isolating Government Officials)	✓	—	✓	—	✓	3

[a]COVID-19: coronavirus disease.

[b]N/A: not applicable.

[c]COVA: Corona Virus Alert.

[d]Absence of information regarding the feature in the app.

[e]HSE: Health Service Executive.

Table 5. Functionality assessment of mobile medical apps (iOS-based only).

No	Name of mobile apps	Knowledge	Tracing/mapping of COVID-19[a] cases	Home monitoring surveillance	Online consultation with health authority	Official mobile app maintained by health authority	Total score
Universal COVID-19 apps							
1	Apple COVID-19	✓	N/A[b]	N/A	N/A	✓	2
2	CDC[c]	✓	N/A	N/A	N/A	✓	2
3	CoronaFACTS	✓	✓	N/A	N/A	✓	3
4	Corona Checker	N/A	N/A	N/A	N/A	N/A	0
5	COVID-19!	✓	✓	N/A	N/A	N/A	2
6	HEALTHLYNKED COVID-19 TRACKER	✓	✓	N/A	N/A	✓	3
7	RELIEF CENTRAL	✓	✓	N/A	N/A	N/A	2
8	Patient Sphere for COVID-19	N/A	N/A	✓	N/A	N/A	1
9	PreMedicus ER	✓	N/A	N/A	✓	N/A	2
10	Mobile Angel Telemed	N/A	N/A	✓	✓	✓	3
Country-specific apps							
11	BMC Combat Covid19	N/A	N/A	✓	✓	✓	3
12	Corona-Care	✓	✓	N/A	N/A	✓	3
13	COVID-19 Gov PK[d]	✓	N/A	N/A	N/A	✓	2
14	Covidom Patient	N/A	N/A	N/A	✓	✓	2
15	COVI QATAR	✓	N/A	N/A	N/A	✓	2
16	COVID-19 UAE[e]	✓	N/A	N/A	N/A	✓	2
17	CUREiTT	✓	N/A	N/A	N/A	N/A	1
18	NJ COVID 19	✓	✓	N/A	N/A	✓	3
19	STOP COVID19 CAT	—[f]	✓	—	—	✓	2
20	Tarassud	✓	✓	N/A	✓	✓	4

[a]COVID-19: coronavirus disease.

[b]N/A: not applicable.

[c]CDC: Centers for Disease Control and Prevention.

[d]PK: Pakistan.

[e]UAE: United Arab Emirates.

[f]Absence of information regarding the feature in the app.

Table 6. Functionality assessment of mobile medical apps (android-based only).

No	Name of mobile apps	Knowledge	Tracing/mapping of COVID-19[a] cases	Home monitoring surveillance	Online consultation with health authority	Official mobile app maintained by health authority	Total score
Universal COVID-19 apps							
1	Test Yourself Goa	✓	N/A[b]	N/A	N/A	✓	2
Country-specific apps							
2	Aarogya Setu	N/A	✓	N/A	N/A	✓	2
3	CoBuddy – Covid19 Tool	N/A	N/A	✓	N/A	N/A	1
4	Corona Watch	✓	✓	N/A	N/A	N/A	2
5	COVI	✓	N/A	N/A	N/A	✓	2
6	Covid-19	N/A	N/A	✓	✓	✓	3
7	COVID-19 NI[c]	✓	N/A	N/A	N/A	✓	2
8	COVID19 Feedback	N/A	N/A	N/A	N/A	N/A	0
9	COVID-19 Quarantine Monitor Tamil Nadu (official)	N/A	N/A	✓	N/A	N/A	1
10	COVID-19 West Bengal Government	N/A	N/A	✓	N/A	N/A	1
11	GCC[d] -Corona Monitoring	N/A	✓	✓	N/A	N/A	2
12	GoK – Direct Kerala	✓	N/A	N/A	N/A	✓	2
13	Home Quarantine (Kwarantanna domowa)	✓	N/A	✓	N/A	✓	3
14	Mahakavach	✓	✓	✓	—[e]	✓	4
15	MP[f] COVID RESPONSE APP	—	—	—	—	—	—
16	Quarantine Watch	—	—	✓	—	✓	2
17	StayHomeSafe	—	—	✓	—	✓	2
18	Test Yourself Puducherry	✓	N/A	N/A	N/A	✓	2

[a]COVID-19: coronavirus disease.

[b]N/A: not applicable.

[c]NI: North Ireland.

[d]GCC: Greater Chennai Corporation.

[e]Absence of information regarding the feature in the app.

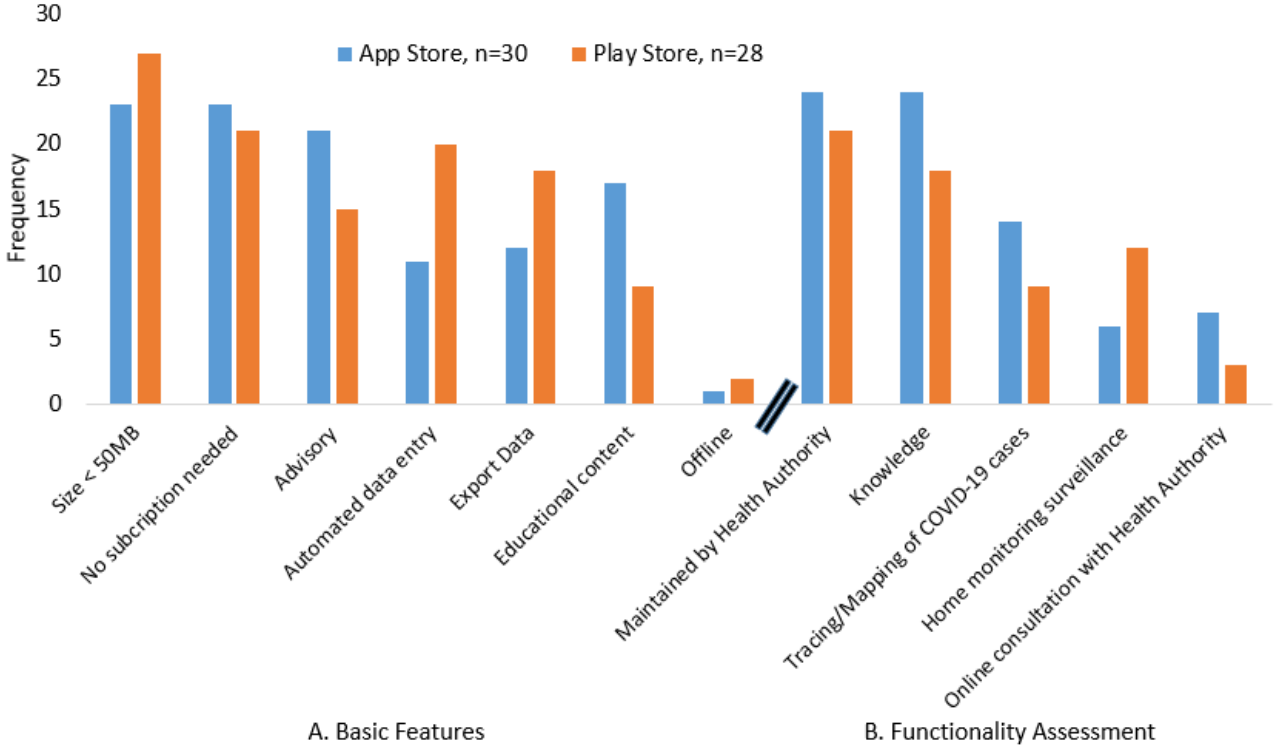
[f]MP: National Health Mission.

The criteria assessed under basic features include the requirement of internet connectivity to use the app, storage capacity, subscription requirement, educational content, ability to export data, automated data entry support, and an advisory feature. According to the results obtained from the assessment of basic features of mobile apps as illustrated in Figure 2A, a majority of the apps from the App Store (29/30, 96.7%) and Google Play (26/28, 92.9%) require internet connectivity to be accessed. There is a higher proportion of mobile apps from Apple (17-23/30, 56.7%-76.7%) that can be accessed without any subscription while providing educational content and advice than Android mobile apps (9-21/28, 32.1%-75.0%). Meanwhile, there are slightly more COVID-19 mobile apps (18-27/28, 64.3%-96.4%) from Google that are less than 50MB in capacity with the ability to export data and allow automated data entry in comparison to Apple (11-23/30, 36.7%-76.7%).

Apart from assessing the basic features, the mobile apps were also assessed based on their functionality as illustrated in Figure 2B. The criteria assessed under functionality includes the availability of COVID-19–related information, tracing or mapping of COVID-19 cases, home monitoring surveillance, online consultation with a health authority, and whether or not the mobile apps are maintained by a health authority. Most of the mHealth apps (7-24/30, 23.3%-80.0%) in the App Store on an iPhone provide better functionality than mHealth apps in the Play Store on an Android smartphone. However, a higher proportion of Android mobile apps (12/28, 42.9%) offer home

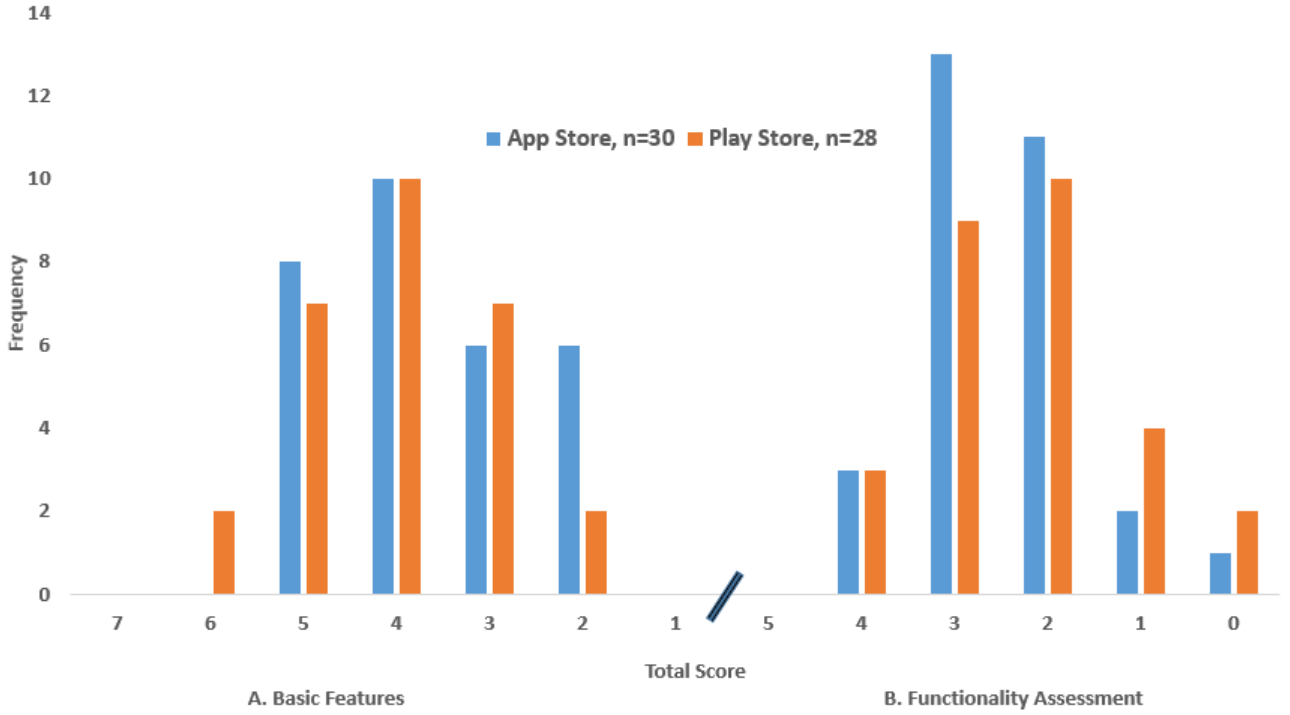
monitoring surveillance related to COVID-19 than Apple (6/30, 20.0%).

Figure 2. Assessment of iOS and android-based mobile apps (A: Basic features; B: Functionality assessment). COVID-19: coronavirus disease.



When assessing the basic features of the mobile apps, none of the apps from the Apple App Store scored 1, 6, or 7, as shown in Figure 3A. Out of the total 30 apps, there were 6 apps (20.0%) that scored two and three, 10 apps (33.3%) that scored four, and 8 apps (26.7%) that scored five. Meanwhile, for apps downloaded from the Google Play Store, none of the apps scored 1 or 7. Out of the total 28 apps, there were 2 apps (7.1%) that scored two, 7 apps (25.0%) that scored three, 10 apps (35.7%) that scored four, 7 apps (25.0%) that scored five, and the remaining 2 apps (7.1%) scored six. In Figure 3B, when assessing the functionality of the mobile apps, none of the apps from the Apple App Store scored 5. There was 1 app (3.3%) that scored zero, 2 apps (6.7%) that scored one, 11 apps (36.7%) scored two, 3 apps (10%) that scored four, and the majority (n=13, 43.3%) scored three. From the Google Play Store, none of the apps scored 5. There were 2 apps (7.1%) that scored zero, 4 apps (14.3%) that scored one, 9 apps (32.1%) that scored three, 3 apps (10.7%) that scored four, and the majority (n=10, 35.7%) scored two.

Figure 3. Total assessment score of iOS and android-based mobile apps (A: Basic features; B: Functionality assessment).



Discussion

Principal Findings

According to the assessment that has been conducted, the mobile apps vary in terms of basic features and functionality. Basic features consist of trivial characteristics that may or may not be of significant importance to users who would like to use COVID-19 mobile apps for COVID-19–related education or self-monitoring purposes with no existing issue for internet access or low mobile storage capacity. Another assessment was conducted based on the advanced features found in the apps. The advanced features touch on the type of content that the app offers on COVID-19, which were used as a measure of the quality of the mobile app. In terms of the basic features, most of the apps from the App Store and the Play Store require no subscriptions and have a storage size of less than 50 MB. There are some apps that need to be paid for, as they need the revenue for advertisements and data mining [23].

People could mistakenly assume that paid mHealth apps have better content or app design. Their value should be assessed in relation to their overall strengths and limitations [21,24]. It is recommended for mHealth apps to have a small storage size, as taking up a lot of phone storage space can result in reduced performance of the mobile phone [25]. Moreover, the App Store offers more apps that provide educational content and advice on COVID-19 than Google Play. This shows the potential of mHealth apps in transforming the delivery of health care services [26]. On the other hand, there are more mobile apps from the Google Play Store that enable data to be exported and offer automated data entry in comparison to the App Store. The availability of data export will allow users to share health reports with their health care providers. Automated data entry can provide greater efficiency in inputting data to the app system while streamlining the user's experience [27,28].

There were 2 COVID-19 mobile apps from that Play Store that had the highest score of six in the assessment of basic features (COVID-19 NI [Northern Ireland] and Test Yourself Puducherry) and 8 COVID-19 mobile apps from the App Store that had the highest score of five (BC COVID-19, COVID Symptom Tracker, Canada COVID-19, Coronavirus Australia, COVID-19 Gov PK [Pakistan], Relief Central, Tarassud, and TraceTogether).

When assessing the functionality of the mobile apps, there were 3 COVID-19 mobile apps from the App Store and the Play Store that had the highest score of four: COVA (coronavirus alert) Punjab and TraceTogether from both the App Store and the Play Store, Tarassud from the App Store, and Mavakavach from the Play Store.

Detailed Review of Mobile Apps in the Apple App Store

Among the apps meant for universal use with an Apple device, APPLE COVID-19 app ranked first in the category of health and fitness with the second-highest user rating of 4.4, and HEALTHLYNKED COVID-19 TRACKER ranked second in the medical category with a slightly higher user rating of 4.6 (rated by thousands of mobile users). Though other apps rated more than 4.5, including CoronaFACTS, Corona Checker, and COVID-19!, there is no strong support that the operating performance of these apps is better than APPLE COVID-19 or HEALTHLYNKED COVID-19 TRACKER because of the low number of users who rated the apps.

APPLE COVID-19 and HEALTHLYNKED COVID-19 TRACKER achieved a score of 4 in the basic feature assessment. One main difference is the ability to access APPLE COVID-19 without requiring any internet or data use, unlike HEALTHYLYNKED COVID-19 TRACKER. However, HEALTHLYNKED COVID-19 TRACKER allows mapping

of COVID-19 cases to provide users with an up-to-date statistic on the number of COVID-19 infections worldwide; a feature that APPLE COVID-19 does not offer. Both apps are run by the health authorities in the United States as APPLE COVID-19 was developed in collaboration with the Centers for Disease Control and Prevention (CDC), while HEALTHLYNKED COVID-19 TRACKER was developed by a company of medical professionals. It is recommended for a developer of health-related apps to collaborate with health care professionals to provide up-to-date content that can be easily trusted by the public [29].

Both APPLE COVID-19 and Relief Central apps scored 2 in the functionality assessment but Relief Central scored an extra point (5) than APPLE COVID-19 (4) in the basic feature assessment, as it supports automated data entry. However, disseminated information in the Relief Central app is not aimed for the general public but those with a background in the medical field, such as health care professionals and students (eg, pharmacy, medical, nursing, and allied health), as the profession of the user is inquired before they can proceed further in using the app. Thus, the content in this app cannot be accessed for the public's perusal.

Both CoronaFACTS and HEALTHLYNKED COVID-19 TRACKER scored the same in both assessments: 4 points in the basic feature assessment and 3 points in the functionality assessment. Both offer the same advanced features of providing information on COVID-19 and mapping of COVID-19 cases, and are maintained by the health authority of their country. The differences between the two apps is that there is no advisory content in CoronaFACTS, unlike HEALTHLYNKED COVID-19 TRACKER, and CoronaFACTS is another app that supports automated data entry, which HEALTHYLYNKED COVID-19 does not support. CoronaFACTS functions by collecting COVID-19–related newspaper articles from trusted sources based on the region chosen by the user, and directs the user to the official site of the article.

Country-specific apps from both stores were also assessed. From the Apple App Store, apps named COVID-19 PK Gov and Tarassud scored the highest (5) in their basic features assessment, with limitations including that users have to be online to operate the app and the absence of a data exportation feature. Other limiting factors of COVID-19 Gov PK in its content are the absence of mapping of COVID-19 cases, home monitoring surveillance, and online consultation with a health authority. Tarassud has better functionality, as it scored 4 out of 5 in the functionality assessment, with only one limitation whereby there is an absence of a home monitoring surveillance feature. Tarassud also includes information on the updated number of cases around the world with guidelines on COVID-19 for those in isolation and quarantine. Of note, using electronically collected influenza data at a Swedish county as an example, the prediction of the influenza virus activity using retrospective surveillance data was successfully integrated into the local health care management system [30].

Detailed Review of Mobile Apps in the Google Play Store

An app named Test Yourself Goa is available for universal use on an Android device. It is an app used to determine if an individual is at high risk of getting COVID-19 based on the symptom-checking feature. After the completion of the symptom-checking test, a piece of advice is given that is aligned with the guidelines proposed by the CDC. For patients who are just learning to manage their disease, mHealth can be of help thanks to its ability to provide advice based on the aggregation of data [31]. In our assessment, it satisfied almost all of the favorable basic features except that it requires an internet connection, has no educational content on the disease, and no automated data entry feature. Test Yourself Goa also allows the health authorities to collect the user's details such as full name, mobile number, zip code, and address. In the functionality assessment, this mobile app only scored 2 out of 5, as it provides information on COVID-19 and is an official app maintained by health authorities.

Meanwhile, among country-specific apps from the Play Store, COVID-19 NI and Test Yourself Puducherry scored the highest in their basic feature assessment with scores of 6 each. The differences between the two apps are that COVID-19 NI does not require internet connectivity, and there is an absence of data exportation in comparison to Test Yourself Puducherry. Both apps also got the same score in their functionality assessment with a considerably low score of 2, proving that they contain COVID-19 information and are maintained by a health authority. The mobile apps can provide disease-specific information to patients, the general public, or health care professionals, which can raise the awareness of users to the signs, causes, and effects of the disease with a view of containing the disease [32].

Detailed Review of Mobile Apps Available in Both Apple App Store and Google Play Store

The only app that is available in both the App Store and the Play Store for universal use is COVID Symptom Tracker. This COVID-19 mobile app enables users to report any symptoms daily. The users need to fill in their personal information such as age, sex, height, weight, and postcode. Additionally, the user's data will be collected by the health authority. Based on the findings in the basic feature assessment, COVID Symptom Tracker requires internet access to operate the app, and it does not provide any educational content. In the functionality assessment, the app scored 3 out of 5 due to the unavailability of several features including home monitoring surveillance and online consultation with a health care professional.

An app named COVA Punjab scored the second-highest in the basic feature assessment with a restriction of the app requiring the internet to work. Most of the mHealth apps in the Apple App Store and the Google Play Store require an internet connection, as it is an important feature for the developers to enable real time data synchronization with the app to prevent the display of outdated information [33,34].

However, the reviewer was not able to access the app, as user registration requires a local phone number. Hence, there is no information about whether or not the app has any educational

content on COVID-19. In the functionality assessment, both COVA Punjab and TraceTogether scored 4 out of 5. TraceTogether does not have a home monitoring surveillance feature, and there is no information on the feature in the COVA Punjab app. However, TraceTogether scored more (5) than COVA Punjab (4) in its basic feature assessment with similar limitations. Although TraceTogether is confirmed to have no educational content, there is a lack of information regarding this feature in COVA Punjab.

TraceTogether has not only the highest score in the basic feature assessment but also has the highest score in the functionality assessment in both the App Store and the Play Store. It is an app used mainly for contact tracing of COVID-19 cases in Singapore to enable users to be notified of people who had close contacts with COVID-19 positive cases. Although contact tracing initially can be done manually, the number of confirmed cases of COVID-19 continues to rise across the world, and it has become difficult to do. Therefore, there are a lot of countries that have used different methods of contact tracing [35]. For example, Taiwan has given health institutions access to patients' travel history and allows relevant authorities to monitor anyone under quarantine by tracking the location of their mobile phone [35]. Meanwhile, in South Korea, the government has maintained a public database of the personal data of patients with COVID-19 including their age, gender, profession, and travel routes [35]. In the case of TraceTogether, if the user is suspected to be infected with COVID-19, the user's data from the app will also be collected by the health authority for contact tracing purposes. TraceTogether also gives information on the functions of the app for a first time user. As the app is being maintained by the health authority, it also allows users to interact with the health authority in addition to allowing users to upload relevant files such as images and documents.

There are also a number of apps that scored low in the functionality assessment. In the Apple App Store, CoronaChecker was identified as a low-scoring app with a score of 0. CoronaChecker is an app intended to provide suggestions on the requirement of a COVID-19 test for the user. Using artificial intelligence, a conversation is initiated with the user upon opening the app. During the interaction, closed questions are generated to ask the user about any clinical symptoms that they may be manifesting to check for any possible viral infection. Despite the low-functionality offered by this app, it was rated 5 out of 5 by 298 users. This could be an indication that the users were satisfied by the operating performance of the app, as CoronaChecker is only designed for users to confirm their health status and not as an educational tool for COVID-19.

COVID19 Feedback obtained the lowest score (0) in the functionality assessment from the Google Play Store. The mobile app only requests feedback from individuals who have done the coronavirus test in India. The feedback received from the public will be used to improve the efficiency and processes related to the coronavirus test. Hence, COVID19 Feedback does not support any of the advanced features related to COVID-19.

In the Apple App Store, CUREiTT scored the lowest in both assessments. It scored two in the basic assessment, as it is only available in exporting data and it supports automated data entry. The app also collects the user's personal information such as phone number, year of birth, and gender. Although CUREiTT only scored one in the functionality assessment, it has a unique function of displaying appropriate clinical trials in the geographical area of individuals diagnosed with cancer or coronavirus.

GoK – Direct Kerala app had the lowest score in both assessments for the Google Play Store. The two scores from the basic feature assessment were due to the storage size of less than 50 MB and the absence of a subscription requirement. However, the app does provide national news reports about the government's actions to manage COVID-19 cases. Moreover, it is an official mobile app maintained by a health authority. Quarantine Watch also has the same score as GoK – Direct Kerala but most of the assessments were not measured, as a national phone number was required for mandatory registration.

Recommendation on COVID-19 Mobile Apps

Every app has a specific goal and a target audience. We have carefully assessed the available apps, which has shown that only a few apps can be used globally for COVID-19 education and self-monitoring. From the previously mentioned results, a high-quality mobile app that can be globally used is COVID Symptom Tracker, which can be installed on both Apple and Android devices. With the aim of slowing the outbreak of the virus, a symptom tracker is useful in identifying high-risk areas in the country, the speed of viral transmission in an area, and individuals who are at the most risk in relation to their health conditions [36].

Based on our study, Mahakavasch and Home Quarantine are the only two apps from the Google Play Store that contain both important features for users: provision of information related to COVID-19 and home-monitoring features. The home-monitoring features are only available for the purpose of monitoring those in quarantine. Home Quarantine, as the name suggests, is designed for those who were assigned to be quarantined at their homes for 14 days. If the users suspect they have been exposed to COVID-19 during their quarantine, they can simply contact the number stated in the app. Mahakavasch has a similar concept as Home Quarantine but it differs slightly, as the users of Mahakavasch will need to take a self-portrait and upload it in the app to allow their location during their quarantine period to be detected.

Self-Monitoring Apps

The use of apps on smartphones to conduct self-assessments for COVID-19 can help to notify the user of their health status. It is also known that technology can promote rapid identification of potential COVID-19 cases for timely interventions to be carried out [37]. This is because COVID-19 is a communicable viral disease in which infected people may appear asymptomatic [36]. With the availability of self-monitoring apps, the user can continue to perform their daily activities at home without going to the hospital for a check-up [21]. Other advantages of self-monitoring apps are the ability to observe a patient's condition at any time, increased efficiency of the health care services with the use of modern technology, and reduced burden

of patients who are immobilized who cannot regularly visit a hospital [38].

Altogether from both stores, there were 6 apps that offer COVID-19 home-monitoring features. The 3 apps from the App Store were BMC Combat COVID19, Patient Sphere for Covid-19, and Mobile Angel. BMC Combat Covid19 is the only app that is specific to citizens of British Columbia, while the other 2 apps are aimed at universal use. BMC Combat Covid19 and Mobile Angel Cancer Telemed scored 2, while Patient Sphere for COVID-19 scored 3 in the basic feature assessment. Patient Sphere for Covid-19 functions similarly as a health record system for a registered individual. Registration can be made through email or phone number. It allows users to manually enter their data in the symptom diary after monitoring themselves. The monitoring parameters that can be recorded are related to common symptoms of COVID-19, such as a stuffy nose, cough, shortness of breath, and chills. However, the Mobile Angel app is restricted to patients from a health care facility that have been registered in the app.

The other 3 apps that offer home-monitoring features are from the Play Store: CoBuddy - Covid19 Tool, COVID-19 Quarantine Monitor Tamil Nadu (official), and COVID-19 West Bengal Government. Unfortunately, the apps are only available for use in specific regions. COVID-19 West Bengal Government and COVID-19 Quarantine Monitor Tamil Nadu (official) scored 4 and CoBuddy - Covid19 Tool scored 3 in basic features. COVID-19 Quarantine Monitor Tamil Nadu (official) and COVID-19 West Bengal Government provide similar contents. A toggle switch should be enabled in the apps to allow the user's daily condition to be monitored.

Quality Improvement of Mobile Apps

Based on the research, there are several recommendations that mobile app developers can consider to improve their existing COVID-19 apps or create a high-quality COVID-19 mobile app in the future.

First, it is recommended that a health-related app be maintained by a health authority to avoid the spread of misleading information to the public. A collaboration with the health authorities to create a mHealth app can increase the reliability of the app, which will encourage more users to be engaged in its use. Otherwise, the user can be informed of the shared information source provided in the app. Second, to increase the engagement rate of the public with the mobile app, it should also contain background information on COVID-19, guidelines, and preventive measures instead of only a focused feature related to COVID-19 (eg, symptom-tracking feature). Moreover, the app should be available for universal use instead of only for residents in a specific country. Third, it is suggested that the apps should be made available without requiring any payment in both the Apple App Store and the Google Play Store to make them more accessible.

Fourth, including real time or near real time updates of statistical analytics with geographical information of positive cases, recovered cases, and a death toll is highly recommended to allow users to be readily informed about the COVID-19 situation worldwide. Fifth, to ensure that users can safely share their personal details, the app should be secure and able to provide assurance to the user that all shared information is kept confidential.

Furthermore, there are other advanced features that can greatly improve the quality of an mobile app including the addition of a feature that can report crowded places to alert users about areas to be avoided to allow them to practice social distancing, as well as a quarantine attendance status, online consultation with health care professionals, and tracing of the whereabouts of positively infected app users to alert their close contacts to undergo contact tracing for COVID-19. It is also crucial to categorize mobile apps into appropriate categories to enable users to find an app easily and thus improve its user engagement rate.

Many COVID-19 mobile apps are appropriately placed under Health and Fitness and Medical. Examples of categories that do not correspond to COVID-19 mobile apps are reference, news, utilities, and tools, as these terms lack specificity in the content they display. Last, an advanced integration of an mHealth app with a health device that can monitor a user's health, such as a digital thermometer that can automatically record the user's body temperature reading in the app, will also enhance the self-monitoring feature of the app.

Limitations

Several limitations were found throughout the study conducted. First, our findings on the available mobile apps in the Google Play Store were limited, as the search for any COVID-19–related keyword has been disabled by Google to avoid any misinformation on the disease. Second, after this research was completed, it is likely that there will be more updated features in the assessed mobile apps. Moreover, new COVID-19 mobile apps may be launched that could not be included in this review. Third, there are gaps in our research, as some apps were inaccessible to the reviewers. This is due to the strict verification process using either a local phone number, especially for apps designed with country-specific functionalities or restricted access for specific users only. There was also 1 app that required payment before the installation of the app could begin. Therefore, some apps were marked with "—" in the assessment results, which indicated the absence of information regarding a certain feature in the app. Fourth, no usability study has been conducted to test for users' responses to COVID-19 mobile apps, so the authors could not conduct a systematic review of literature regarding COVID-19 mobile apps.

Conclusions

It is important to evaluate the contents and features of COVID-19 mobile apps to guide users in choosing a suitable mobile app based on their requirements and help developers to improve the designs of their existing or future mobile apps to further enhance quality. Evaluation of basic functions showed that 75.0% (n=36) of the included 48 mobile apps do not require a subscription, 56.3% (n=27) provide symptom advice, and 41.7% (n=20) have educational content. In terms of the specific functions, more than half of the included mobile apps are official mobile apps maintained by a health authority for COVID-19 information provision. Around 37.5% (n=18) and 31.3% (n=15)

of the mobile apps have tracing or mapping and home monitoring surveillance functions, respectively, with only 17% (n=8) of mobile apps equipped with an online consultation function. Quality-wise, 58.3% (n=28) of the included mobile apps scored 4 points and above out of the maximum 7, proving that during the time constraint of a few weeks, the mobile app developers did not manage to create a fully comprehensive mobile app. Our study paves the way for future work to determine the role of mobile apps in controlling the rate of COVID-19 transmission.

Authors' Contributions

LCM, NU, NAA, NO, NK, and HPG contributed to the study design, data extraction, functionality assessment, analysis and interpretation of data, and drafting of the manuscript. CST, KWG, PWN, YMA-W, and KSL contributed to the functionality assessment and analysis, interpretation of data, and revising of the article. All authors proofread and approved the submitted version of the article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristic of mobile medical apps (iOS-based).
[DOCX File , 27862 KB - mhealth_v8i9e19796_app1.docx]

Multimedia Appendix 2
Characteristic of mobile medical apps (android-based).
[PDF File (Adobe PDF File), 574 KB - mhealth_v8i9e19796_app2.pdf]

References

1. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected: interim guidance. World Health Organization. 2020 Mar 13. URL: https://apps.who.int/iris/bitstream/handle/10665/331446/WHO-2019-nCoV-clinical-2020.4-eng.pdf?sequence=1&isAllowed=y [accessed 2020-04-22]
2. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
3. Chen C, Jyan H, Chien S, Jen H, Hsu C, Lee P, et al. Containing COVID-19 among 627,386 persons in contact with the Diamond Princess cruise ship passengers who disembarked in Taiwan: big data analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]
4. Materia FT, Faasse K, Smyth JM. Understanding and preventing health concerns about emerging mobile health technologies. JMIR Mhealth Uhealth 2020 May 25;8(5):e14375 [FREE Full text] [doi: 10.2196/14375] [Medline: 32449688]
5. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
6. Hilty D, Chan S, Torous J, Luo J, Boland R. A framework for competencies for the use of mobile technologies in psychiatry and medicine: scoping review. JMIR Mhealth Uhealth 2020 Feb 21;8(2):e12229 [FREE Full text] [doi: 10.2196/12229] [Medline: 32130153]
7. Lee S, Lee Y, Lee S, Islam SMS, Kim S. Toward developing a standardized core set of outcome measures in mobile health interventions for tuberculosis management: systematic review. JMIR Mhealth Uhealth 2019 Feb 19;7(2):e12385 [FREE Full text] [doi: 10.2196/12385] [Medline: 30777847]
8. 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]
9. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
10. Apidi NA, Murugiah MK, Muthuveloo R, Soh YC, Caruso V, Patel R, et al. Mobile medical applications for dosage recommendation, drug adverse reaction, and drug interaction: review and comparison. Ther Innov Regul Sci 2017 Jul;51(4):480-485. [doi: 10.1177/2168479017696266] [Medline: 30227053]
11. Izahar S, Lean QY, Hameed MA, Murugiah MK, Patel RP, Al-Worafi YM, et al. Content analysis of mobile health applications on diabetes mellitus. Front Endocrinol (Lausanne) 2017;8:318. [doi: 10.3389/fendo.2017.00318] [Medline: 29230195]

12. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. Nature 2019 Feb;566(7745):467-474 [FREE Full text] [doi: 10.1038/s41586-019-0956-2] [Medline: 30814711]
13. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
14. Were MC, Sinha C, Catalani C. A systematic approach to equity assessment for digital health interventions: case example of mobile personal health records. J Am Med Inform Assoc 2019 Aug 01;26(8-9):884-890. [doi: 10.1093/jamia/ocz071] [Medline: 31188438]
15. Ossemane EB, Moon TD, Were MC, Heitman E. Ethical issues in the use of SMS messaging in HIV care and treatment in low- and middle-income countries: case examples from Mozambique. J Am Med Inform Assoc 2018 Apr 01;25(4):423-427 [FREE Full text] [doi: 10.1093/jamia/ocx123] [Medline: 29088384]
16. Morse SS, Murugiah MK, Soh YC, Wong TW, Ming LC. Mobile health applications for pediatric care: review and comparison. Ther Innov Regul Sci 2018 May;52(3):383-391. [doi: 10.1177/2168479017725557] [Medline: 29714538]
17. Device software functions including mobile medical applications. US Food and Drug Administration. 2019. URL: https://www.fda.gov/medical-devices/digital-health/device-software-functions-including-mobile-medical-applications [accessed 2020-04-22]
18. Kessel KA, Vogel MM, Schmidt-Graf F, Combs SE. Mobile apps in oncology: a survey on health care professionals' attitude toward telemedicine, mHealth, and oncological apps. J Med Internet Res 2016 Nov 24;18(11):e312 [FREE Full text] [doi: 10.2196/jmir.6399] [Medline: 27884810]
19. Andriesen J, Bull S, Dietrich J, Haberer JE, Van Der Pol B, Voronin Y, et al. Using digital technologies in clinical HIV research: real-world applications and considerations for future work. J Med Internet Res 2017 Jul 31;19(7):e274 [FREE Full text] [doi: 10.2196/jmir.7513] [Medline: 28760729]
20. Vaggers S, Puri P, Wagenlehner F, Somani B. A content analysis of mobile phone apps for the diagnosis, treatment, and prevention of urinary tract infections, and their compliance with European Association of urology guidelines on urological infections. Eur Urology Focus 2020. [doi: 10.1016/j.euf.2020.02.002]
21. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
22. Anthony Berauk VL, Murugiah MK, Soh YC, Chuan Sheng Y, Wong TW, Ming LC. Mobile health applications for caring of older people: review and comparison. Ther Innov Regul Sci 2018 May;52(3):374-382. [doi: 10.1177/2168479017725556] [Medline: 29714532]
23. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
24. O'Connor Y, Andreev P, O'Reilly P. MHealth and perceived quality of care delivery: a conceptual model and validation. BMC Med Inform Decis Mak 2020 Feb 27;20(1):41 [FREE Full text] [doi: 10.1186/s12911-020-1049-8] [Medline: 32103746]
25. Aljaber T, Gordon N, editors. A Hybrid Evaluation Approach and Guidance for mHealth Education Apps. New York City: Springer; 2017.
26. Lavallee DC, Lee JR, Austin E, Bloch R, Lawrence SO, McCall D, et al. mHealth and patient generated health data: stakeholder perspectives on opportunities and barriers for transforming healthcare. Mhealth 2020;6:8. [doi: 10.21037/mhealth.2019.09.17] [Medline: 32190619]
27. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile apps for blood pressure monitoring: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
28. Kim JY, Wineinger NE, Taitel M, Radin JM, Akinbosoye O, Jiang J, et al. Self-monitoring utilization patterns among individuals in an incentivized program for healthy behaviors. J Med Internet Res 2016 Nov 17;18(11):e292 [FREE Full text] [doi: 10.2196/jmir.6371] [Medline: 27856407]
29. Krishnan G, Selvam G. Factors influencing the download of mobile health apps: content review-led regression analysis. Health Policy Technol 2019 Dec;8(4):356-364. [doi: 10.1016/j.hlpt.2019.09.001]
30. Spreco A, Eriksson O, Dahlström Ö, Cowling BJ, Timpka T. Integrated detection and prediction of influenza activity for real-time surveillance: algorithm design. J Med Internet Res 2017 Jun 15;19(6):e211 [FREE Full text] [doi: 10.2196/jmir.7101] [Medline: 28619700]
31. Morley J, Floridi L. The limits of empowerment: how to reframe the role of mHealth tools in the healthcare ecosystem. Sci Eng Ethics 2020 Jun;26(3):1159-1183 [FREE Full text] [doi: 10.1007/s11948-019-00115-1] [Medline: 31172424]
32. Otu A, Ebenso B, Okuzu O, Osifo-Dawodu E. Using a mHealth tutorial application to change knowledge and attitude of frontline health workers to Ebola virus disease in Nigeria: a before-and-after study. Hum Resour Health 2016 Feb 12;14:5 [FREE Full text] [doi: 10.1186/s12960-016-0100-4] [Medline: 26872824]

33. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011 Sep 22;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
34. Kollmann A, Riedl M, Kastner P, Schreier G, Ludvik B. Feasibility of a mobile phone-based data service for functional insulin treatment of type 1 diabetes mellitus patients. J Med Internet Res 2007 Dec 31;9(5):e36 [FREE Full text] [doi: 10.2196/jmir.9.5.e36] [Medline: 18166525]
35. Cho H, Ippolito D, Yu Y. Contact tracing mobile apps for covid-19: Privacy considerations and related trade-offs [preprint]. arXiv 2020.
36. Mayor S. Covid-19: Researchers launch app to track spread of symptoms in the UK. BMJ 2020 Mar 27;368:m1263. [doi: 10.1136/bmj.m1263] [Medline: 32220898]
37. Srinivasa Rao ASR, Vazquez JA. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine. Infect Control Hosp Epidemiol 2020 Mar 03:1-5 [FREE Full text] [doi: 10.1017/ice.2020.61] [Medline: 32122430]
38. Malasinghe LP, Ramzan N, Dahal K. Remote patient monitoring: a comprehensive study. J Ambient Intell Human Comput 2017 Oct 26;10(1):57-76. [doi: 10.1007/s12652-017-0598-x]

Abbreviations

CDC: Centers for Disease Control and Prevention
COVA: coronavirus alert
COVID-19: coronavirus disease
mHealth: mobile health
NI: Northern Ireland
PK: Pakistan
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
VPN: virtual private network

Edited by G Eysenbach; submitted 01.05.20; peer-reviewed by T Khan, J Santos, C Liao; comments to author 09.06.20; revised version received 23.06.20; accepted 30.06.20; published 16.09.20.

Please cite as:
Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, Goh KW, Ng PW, Al-Worafi YM, Lee KS, Goh HP
Mobile Health Apps on COVID-19 Launched in the Early Days of the Pandemic: Content Analysis and Review
JMIR Mhealth Uhealth 2020;8(9):e19796
URL: https://mhealth.jmir.org/2020/9/e19796
doi:10.2196/19796
PMID:32609622

Original Paper

Communication Technology Preferences of Hospitalized and Institutionalized Frail Older Adults During COVID-19 Confinement: Cross-Sectional Survey Study

Guillaume Sacco[1,2,3], MD, PhD; Sébastien Lléonart[4], MD; Romain Simon[1,3], MS; Frédéric Noublanche[1,2,3], MS; Cédric Annweiler[1,2,3,5], MD, PhD; TOVID Study Group[1,4,6]

[1]Department of Geriatric Medicine and Memory Clinic, University Hospital of Angers, Angers, France

[2]Laboratoire de Psychologie des Pays de la Loire (LPPL - EA 4638), SFR Confluences, Université Nantes-Angers-Le Mans, Nantes, France

[3]ALLEGRO Living Lab Research Center on Autonomy and Longevity, University Hospital of Angers, Angers, France

[4]Pôle médico-social Saint Nicolas, Univeristy Hospital of Angers, Angers, France

[5]Robarts Research Institute, Department of Medical Biophysics, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada

[6]School of Medicine, Health Faculty, University of Angers, Angers, France

Corresponding Author:
Guillaume Sacco, MD, PhD
Department of Geriatric Medicine and Memory Clinic
University Hospital of Angers
4 Rue Larey
Angers
France
Phone: 33 609733254
Email: yogisacco@gmail.com

Abstract

Background: Technological communication methods such as telephone calls and video calls can help prevent social isolation and loneliness in frail older adults during confinement.

Objective: Our objectives were to determine which virtual communication method (ie, telephone call or video call) was preferred by confined older hospital patients and nursing home residents and the variables influencing this preference.

Methods: The TOVID (Telephony Or Videophony for Isolated elDerly) study was a cross-sectional study that was designed to examine the preference between telephone calls and video calls among frail older adults who were either hospitalized in a geriatric acute care unit or institutionalized in a long-term care and nursing home during the COVID-19 confinement period.

Results: A total of 132 older people were surveyed between March 25 and May 11, 2020 (mean age 88.2 years, SD 6.2); 79 (59.8%) were women. Patients hospitalized in the geriatric acute care unit were more able to establish communication independently than residents institutionalized in the long-term care and nursing home (P=.03) and were more satisfied with their communication experiences (P=.02). Overall, older people tended to favor telephone calls (73/132, 55.3%) over video calls (59/132, 44.7%); however, their satisfaction degree was similar regardless of the chosen method (P=.1), with no effect of age (P=.97) or gender (P=.2). In the geriatric acute care unit, the satisfaction degrees were similar for telephone calls (40/41, 98%) and video calls (33/38, 87%) in older patients (P=.10). Conversely, in the long-term care and nursing home, residents were more satisfied with the use of video calls to communicate with their relatives (14/15, 93%) versus the use of telephone calls (6/12, 50%; P=.02).

Conclusions: Older people confined to health care settings were able to complete telephone calls more independently than video calls, and they tended to use telephone calls more often than video calls. The satisfaction degrees were similar with both modalities and even greater with video calls among long-term care and nursing home residents when they were given assistance to establish communication.

Trial Registration: ClinicalTrials.gov NCT04333849: https://www.clinicaltrials.gov/ct2/show/NCT04333849.

(JMIR Mhealth Uhealth 2020;8(9):e21845) doi:10.2196/21845

KEYWORDS

video communication; telephone; older adults; nursing home; hospital; confinement; elderly; COVID-19; communication; technology; social isolation; loneliness

Introduction

All serious epidemics, such as the COVID-19 pandemic, prompt social organizations to deeply rethink the services they offer, especially toward frail older adults. For instance, the visitation restrictions in geriatric care units and nursing homes, although essential to the control of the pandemic, have also become a major source of social isolation and loneliness for vulnerable populations [1-3]. The only remaining social link for patients and residents in these facilities during confinement has been virtual communication using a range of communication devices, notably telephone calls and video calls.

The use of video calls is currently increasing because of the innovative nature of video, which allows people to speak to and hear others, view their expressions, and establish richer relationships than may be possible with a simple telephone call [4]. The research question in this paper is whether video calling meets the real demands and expectations of older adults or whether they prefer more traditional communication methods, such as the telephone**,** to contact their relatives. Moreover, evidence from previous literature about the efficacy of electronic interventions to avoid social isolation among older adults is weak and inconsistent [5-7].

Like many other practitioners, from the start of the COVID-19 public health crisis and the first visit restrictions for hospitalized and institutionalized patients, we proposed to organize daily communications between the patients and their relatives to avoid excessive isolation [8,9]. Both telephone calls and video calls were offered to the patients. The main objective of this study was to determine which virtual communication mode was preferred by older patients confined in hospital or institutionalized in a nursing home. Our secondary aims were to identify the proportions of older patients and residents who could independently communicate with virtual support; to measure and compare the degrees of satisfaction following telephone calls and video calls; and to analyze the effects of age and context (hospitalization versus nursing home) on the communication mode choice of older adults.

Methods

Design and Settings

The cross-sectional TOVID (Telephony Or Videophony for Isolated elDerly) study was conducted in the geriatric acute care unit and in the long-term care unit and nursing home of the University Hospital of Angers, France, between March 25 and May 11, 2020, during the national confinement period in France. No outside visits to these hospital units were authorized during this period. The study was conducted in accordance with the ethical standards set forth in the Helsinki Declaration (1983) and was approved by the local ethics committee (number 2020/29). The study protocol was declared to the National Commission for Information Technology and Civil Liberties (CNIL) under the number ar20-0030v0 (ClinicalTrials.gov identifier: NCT04333849).

Participants

All older adults consecutively hospitalized in the geriatric acute care unit and in the long-term care and nursing home were considered for inclusion in the study. Patients who refused to participate or who were unable to communicate with their relatives for medical reasons were not included in the study.

Data Collection

Health professionals accustomed to using communication devices visited all eligible patients at least once per day to offer to help them organize their communications with their relatives. All patients who expressed interest were offered either a telephone call or a video call, and they were clearly informed that they could receive assistance to establish communication if necessary. All cognitively intact patients who objected to any help were considered to be independently capable of establishing communication (ie, declarative measure). The details of the communication (application and equipment, schedule, and duration) were discussed with the relatives prior to the communication to ensure that the call proceeded smoothly and easily.

Data regarding the participants' age, gender, hospital unit (ie, geriatric acute care unit or long-term care and nursing home), independent ability to establish communication, preferred virtual communication mode, and degree of satisfaction toward the mode of communication were collected. The preferred virtual communication mode was identified using a standardized question with three options (nothing, telephone call, or video call). The satisfaction degree toward virtual communication was assessed after the communication using a 6-point Likert scale (with 1=not satisfied at all to 6=totally satisfied), and the proportion of people satisfied with the communication (defined as a score ≥5/6) was analyzed.

Number of Participants

Because the main objective of the study was descriptive, it was not necessary to calculate the number of subjects. However, to match a normal distribution of the data and to use parametric statistical tests, at least 30 participants were required in each group (ie, n=30 in the telephone call group and n=30 in the video call group).

Statistical Analyses

Categorical variables were described using numbers and percentages and quantitative variables were described using means and standard deviations, as appropriate. Comparisons between older patients in the geriatric acute care unit and the long-term care and nursing home and between patients who chose telephone calls and video calls were performed using chi-square tests for qualitative variables (or the exact Fisher test where appropriate), and the Student *t* test was used for quantitative variables (or the Mann-Whitney U test where

appropriate). *P* values <.05 were considered significant. All statistical analyses were performed with SPSS version 20 (IBM Corporation).

Availability of Data and Materials

Patient level data are freely available from the co-corresponding author at Cedric.Annweiler@chu-angers.fr after notification and authorization of the competent authorities. There is no personal identification risk with these anonymized raw data.

Results

We invited 163 older adults to take part in the study; 132 (80.1%) agreed to participate and were included in the study. The age range of the 132 participants was 66 to 103 years (mean age 88.2 years, SD 6.2); 78 (59.1%) were women.

As illustrated in Table 1, the participants tended to favor telephone calls (73/132, 55.3%) over video calls (59/132, 44.7%). The satisfaction degrees with the two modalities were similar (46/73, 87%, with telephone calls versus 47/59, 89%, with video calls, *P*=.10). There was no effect of age (*P*=.97) or gender (*P*=.16) on the choice of virtual communication mode.

Patients hospitalized in the geriatric acute care unit were more frequently able to independently establish communication (24/105, 22.8%) than residents institutionalized in the long-term care and nursing home (1/27, 3.8%, *P*=.03) and they were more able to independently establish communication by telephone call (22/73, 30.1%) than by video call (3/59, 5.1%; *P*<.001). Moreover, patients hospitalized in the geriatric acute care unit were more often satisfied with the communication (73/79, 92%) than residents of the long-term care and nursing home (20/27, 74%; *P*=.02).

Table 1. Characteristics of the study participants (N=132).

Characteristic	Whole cohort	Preferred virtual communication mode			Location of confinement		
	(N=132)	Telephone call (n=73)	Video call (n=59)	*P* value[a]	Geriatric acute care unit (n=105)	Long-term care and nursing home (n=27)	*P* value[a]
Sociodemographic data							
Age (years), mean (SD)	88.2 (6.2)	88.2 (5.7)	88.2 (6.7)	.97	88.8 (5.4)	85.8 (8.1)	.08
Female gender, n (%)	78 (59.1)	39 (53.4)	39 (66.1)	.16	57 (54.3)	21 (77.8)	.03
Communication with relatives, n (%)							
Capability of independently establishing communication[b]	25 (19.1)	22 (30.1)	3 (5.2)	<.001	24 (22.8)	1 (3.8)	.03
Choice of telephone call	73 (55.3)	N/A[c]	N/A	N/A	61 (58.1)	12 (44.4)	.28
High degree of satisfaction (Likert scale score ≥5/6)[d]	93 (87.7)	46 (86.8)	47 (88.7)	>.99	73 (92.4)	20 (74.1)	.02

[a]Comparisons based on chi-square test or exact Fisher test for qualitative variables and Student *t* test or Mann-Whitney U test for quantitative variables, as appropriate.

[b]Data missing for 1 participant in the long-term care and nursing home.

[c]N/A: not applicable.

[d]Data missing for 26 participants in the geriatric acute care unit.

In the subgroup analyses (Table 2), older patients chose telephone calls and video calls at similar frequencies when they were hospitalized in the geriatric acute care unit (61/106, 57.5% and 44/106, 41.5%, respectively) or institutionalized in the long-term care and nursing home (12/27, 44.4%, and 12/27, 55.6%, respectively). In the geriatric acute care unit, the satisfaction degrees were similar for telephone calls (40/41, 98%) and video calls (33/38, 87%) in older patients (*P*=.10). Conversely, in the long-term care and nursing home, residents were more often satisfied with the use of video calls to communicate with their relatives (14/15, 93%, versus 6/12, 50%, *P*=.02) (Table 2).

Table 2. Subgroup analyses according to confinement place (N=132).

Characteristic	Patients in the geriatric acute care unit			Residents in the long-term care and nursing home		
	Choice of telephone call (n=61)	Choice of video call (n=44)	*P* value[a]	Choice of telephone call (n=12)	Choice of video call (n=15)	*P* value[a]
Sociodemographic data						
Age (years), mean (SD)	88.5 (5.6)	89.3 (5.2)	.49	86.8 (6.1)	85.0 (9.6)	.59
Female gender, n (%)	31 (51)	26 (59)	.43	8 (67)	13 (87)	.36
Communication with relatives, n (%)						
Capability of independently establishing communication[b]	22 (36)	2 (5)	<.001	0 (0)	1 (7)	>.99
High degree of satisfaction (≥5/6)[c]	40 (98)	33 (87)	.10	6 (50)	14 (93)	.02

[a]Comparisons based on chi-square test or exact Fisher test for qualitative variables and Student *t* test or Mann-Whitney U test for quantitative variables, as appropriate.

[b]Data missing for 1 participant in the long-term care and nursing home.

[c] Data missing for 26 participants in the geriatric acute care unit.

Discussion

Principal Results

We found that older adults confined to health care settings (ie, a geriatric acute care unit or long-term care and nursing home) were more often independently able to perform telephone calls than video calls, and they tended to use the telephone more often to communicate with their relatives. Their levels of satisfaction were similar with both communication supports, and satisfaction was even greater with video calls among residents of the long-term care and nursing home when they received assistance to establish communication.

Limitations

Our study has some limitations. First, the study was monocentric, which limits the representativeness of the study population even if we were able to include a relatively high number of participants. Second, the results should be interpreted with caution because some confounding factors such as cognition and mood were not assessed. Larger, and if possible prospective, studies should be conducted on different population groups to better understand the need for video calls and their effects on loneliness, social isolation, and quality of life in older adults.

Comparison With Prior Work

The global confinement during the COVID-19 pandemic has highlighted the importance of preventing social isolation and loneliness in older adults [1,2], as social disconnection is associated with increased anxiety and depression [10] and loneliness is associated with increased risks of health disorders, including major neurocognitive disorders [11]. Access to social technologies has been widely proposed to reduce social isolation during the pandemic [8,12], although uncertainty remains about efficacy of video call interventions to reduce loneliness in older adults [13]. Despite this, studies focusing on these technological interventions among frail older adults remain rare [14], and to our knowledge, no study has evaluated the acceptance and preferences of older adults regarding these interventions. However, there is some evidence that a weak digital culture and eventual impairments may complicate the use of virtual communication modes by older adults [15].

Technological acceptance is a balance between perceived usefulness and perceived simplicity of use [16], to which is added the quality of the output (ie, the way in which the system performs the expected task [17]). A model of technology acceptance dedicated to older adults brings together 10 factors influencing the acceptance of the technology: the value (utility), usability (perceived simplicity), affordability, accessibility, technical support, social support, emotion (output quality), independence, experience, and confidence [15]. In our study, pre-experience (ie, perceived utility and perceived simplicity of use, reflected by the choice of communication mode) and post-experience perceptions (ie, output quality, reflected by the satisfaction degree) were different between geriatric acute care unit patients and long-term care and nursing home residents. In our study, the preference for telephone calls (with 55% of participants making this choice) is likely due to the participants' pre-experience of the perceived simplicity of this device compared to video calls, which are more often misunderstood and less often used by people in this older generation. In addition, the postexperience perception differed between geriatric acute care unit patients and long-term care and nursing home residents, with a higher degree of satisfaction in the geriatric acute care unit than in the long-term care and nursing home (73/29, 92%, vs 20/27, 74%; *P*=.02). One possible explanation is that the proportion of older adults who were unable to establish communication by themselves was higher in the long-term care and nursing home (25/26, 96%) than in the geriatric acute care unit (81/105, 77.1%; *P*=.03). It therefore appears that the pre-experience was unbalanced by the importance given to the perceived ease of use compared to the perceived usefulness of adding video to the call. On the other hand, when older adults in the long-term care and nursing home chose the video calls, generally with assistance establishing the communication, their satisfaction degree was higher with the video calls, which shows that their post-experience was modified

and could possibly modify their future choices and communication habits.

We also found that a greater proportion of older adults needed assistance to use video calls (56/59, 94.8%) than to use telephone calls (51/73, 69.9%, $P<.001$). Previous literature has emphasized the importance of offering accessible communication systems, and sustained efforts should be pursued to simplify communication technologies, particularly video calls [14,18]. It would also be interesting to support or adapt these communication technologies for people who have communication issues. Technology may provide support for this.

Conclusions

Older adults in a geriatric acute care unit and a long-term care and nursing home were more independently able to make telephone calls than video calls, and they tended to use the telephone more often than video. However, their post-experience satisfaction with video calls was high. This cross-sectional study contributes to understanding the acceptance and the challenges of frail older adults regarding video calls.

Acknowledgments

GS is supported by a postdoctoral grant from Research Center on Autonomy and Longevity, University Hospital of Angers, France (2019-2020). The sponsors had no role in the design and conduct of the study, in the collection, management, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript. The authors wish to thank Tikeasy and the Fondation Hôpitaux de Paris Hôpitaux de France for the loan of the tablets for the video calls. The authors also thank the TOVID study group: Françoise Angebault (Saint Nicolas University Hospital, Angers, France), Cédric Annweiler (Department of Geriatric Medicine, University Hospital, and School of Medicine, Health Faculty, University of Angers, Angers, France), Marine Asfar (Department of Geriatric Medicine, University Hospital, Angers, France), Melinda Beaudenon (Department of Geriatric Medicine, University Hospital, Angers, France), France Bertrand (School of Medicine, Health Faculty, University of Angers, Angers, France), Marie Brichard (School of Medicine, Health Faculty, University of Angers, Angers, France), Jean Barré (Department of Geriatric Medicine, University Hospital, Angers, France), Gabriel Bordier (School of Medicine, Health Faculty, University of Angers, Angers, France), Inès Bouchet (School of Medicine, Health Faculty, University of Angers, Angers, France), Aude Bouchevreau (School of Medicine, Health Faculty, University of Angers, Angers, France), Antoine Brangier (Department of Geriatric Medicine, University Hospital, Angers, France), Adeline Chanson (School of Medicine, Health Faculty, University of Angers, Angers, France), Isabelle Chausson (Saint Nicolas University Hospital, Angers, France) , Linda Collet (School of Medicine, Health Faculty, University of Angers, Angers, France), Marie Daussy (School of Medicine, Health Faculty, University of Angers, Angers, France), Luc Delneste (School of Medicine, Health Faculty, University of Angers, Angers, France), Elise Dolbeau (School of Medicine, Health Faculty, University of Angers, Angers, France), Maxime Dronne (School of Medicine, Health Faculty, University of Angers, Angers, France), Brune Du Plessis (School of Medicine, Health Faculty, University of Angers, Angers, France), Guillaume Duval (Department of Geriatric Medicine, University Hospital, Angers, France), Lydia Gabard (School of Medicine, Health Faculty, University of Angers, Angers, France), Jennifer Gautier (Department of Geriatric Medicine, University Hospital, Angers, France), Anne Guyard (Department of Geriatric Medicine, University Hospital, Angers, France), Tiphaine Housty (Department of Geriatric Medicine, University Hospital, Angers, France), Valentin Jouanno (School of Medicine, Health Faculty, University of Angers, Angers, France), Clémence Le Ho (School of Medicine, Health Faculty, University of Angers, Angers, France), Cassandra Lepoureau (School of Medicine, Health Faculty, University of Angers, Angers, France), Sébastien Lléonart, Mélanie Loison (School of Medicine, Health Faculty, University of Angers, Angers, France), Jocelyne Loison (Department of Geriatric Medicine, University Hospital, Angers, France), Laurence Loufrani (Saint Nicolas University Hospital, Angers, France), Emilie Lucas (Department of Geriatric Medicine, University Hospital, Angers, France), Camille Macé (School of Medicine, Health Faculty, University of Angers, Angers, France), Naomi Maire (School of Medicine, Health Faculty, University of Angers, Angers, France), Clément Marla (School of Medicine, Health Faculty, University of Angers, Angers, France), Nathalie Martin (Saint Nicolas University Hospital, Angers, France), Andréa Mathieu (School of Medicine, Health Faculty, University of Angers, Angers, France), Morgane Mauger (School of Medicine, Health Faculty, University of Angers, Angers, France), Sahondra Mayet (School of Medicine, Health Faculty, University of Angers, Angers, France), Justine Menon (School of Medicine, Health Faculty, University of Angers, Angers, France), Frédéric Messian (School of Medicine, Health Faculty, University of Angers, Angers, France), Constance Moucheboeuf (School of Medicine, Health Faculty, University of Angers, Angers, France), Antoine Mussard (School of Medicine, Health Faculty, University of Angers, Angers, France), Frédéric Noublanche (Department of Geriatric Medicine, University Hospital, Angers, France), Antoine Octau (School of Medicine, Health Faculty, University of Angers, Angers, France), Marie Otekpo (Department of Geriatric Medicine, University Hospital, Angers, France), Capucine Phuong-Anh Carion (School of Medicine, Health Faculty, University of Angers, Angers, France), Valentin Poile (School of Medicine, Health Faculty, University of Angers, Angers, France), Lisa Quentin (School of Medicine, Health Faculty, University of Angers, Angers, France), Anne Renaudin (Department of Geriatric Medicine, University Hospital, Angers, France), Gwladys Roger (School of Medicine, Health Faculty, University of Angers, Angers, France), Quentin Rousseau (School of Medicine, Health Faculty, University of Angers, Angers, France), Florence Roult (Department of Geriatric Medicine, University Hospital, Angers, France), Guillaume Sacco (Department of Geriatric Medicine, University Hospital, Angers, France), Claire Simon (School of Medicine, Health Faculty, University of Angers, Angers, France), Ophélie Simon (School of Medicine, Health

Faculty, University of Angers, Angers, France), Romain Simon (Department of Geriatric Medicine, University Hospital, Angers, France), Romane Surget (School of Medicine, Health Faculty, University of Angers, Angers, France), Violette Tourrel (School of Medicine, Health Faculty, University of Angers, Angers, France), Lise Trubert (School of Medicine, Health Faculty, University of Angers, Angers, France), Inès Watteaux (School of Medicine, Health Faculty, University of Angers, Angers, France).

Authors' Contributions

CA has full access to all of the data in the study, takes responsibility for the data, the analyses and interpretation, and has the right to publish any and all data, separate and apart from the attitudes of the sponsors. CA supervised the study and provided administrative, technical, and material support. GS and CA conceived and designed the study, analyzed and interpreted the data, drafted the manuscript, and provided statistical expertise. SL and RS provided critical revision of the manuscript for important intellectual content. All authors have read and approved the manuscript.

Conflicts of Interest

None declared.

References

1. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health 2020 May;5(5):e256. [doi: 10.1016/s2468-2667(20)30061-x]
2. Cudjoe TK, Kotwal AA. "Social Distancing" Amid a Crisis in Social Isolation and Loneliness. J Am Geriatr Soc 2020 Jun 15;68(6):E27-E29 [FREE Full text] [doi: 10.1111/jgs.16527] [Medline: 32359072]
3. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
4. Hemberg J, Santamäki Fischer R. A Window Toward the World Older Adults' Experiences of Becoming in Health and Developing as Human Beings Through Interacting With Others Using Real Video Communication. Holist Nurs Pract 2018;32(2):90-97. [doi: 10.1097/hnp.0000000000000254]
5. Chipps J, Jarvis MA, Ramlall S. The effectiveness of e-Interventions on reducing social isolation in older persons: A systematic review of systematic reviews. J Telemed Telecare 2017 Dec;23(10):817-827. [doi: 10.1177/1357633X17733773] [Medline: 28958209]
6. Gardiner C, Geldenhuys G, Gott M. Interventions to reduce social isolation and loneliness among older people: an integrative review. Health Soc Care Community 2018 Mar;26(2):147-157. [doi: 10.1111/hsc.12367] [Medline: 27413007]
7. Poscia A, Stojanovic J, La Milia DI, Duplaga M, Grysztar M, Moscato U, et al. Interventions targeting loneliness and social isolation among the older people: An update systematic review. Exp Gerontol 2018 Feb;102:133-144. [doi: 10.1016/j.exger.2017.11.017] [Medline: 29199121]
8. Eghtesadi M. Breaking Social Isolation Amidst COVID-19: A Viewpoint on Improving Access to Technology in Long-Term Care Facilities. J Am Geriatr Soc 2020 May;68(5):949-950 [FREE Full text] [doi: 10.1111/jgs.16478] [Medline: 32277470]
9. Ruopp MD. Overcoming the Challenge of Family Separation From Nursing Home Residents During COVID-19. J Am Med Dir Assoc 2020 Jul;21(7):984-985 [FREE Full text] [doi: 10.1016/j.jamda.2020.05.022] [Medline: 32536554]
10. Santini ZI, Jose PE, York Cornwell E, Koyanagi A, Nielsen L, Hinrichsen C, et al. Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. Lancet Public Health 2020 Jan;5(1):e62-e70 [FREE Full text] [doi: 10.1016/S2468-2667(19)30230-0] [Medline: 31910981]
11. Sutin A, Stephan Y, Luchetti M, Terracciano A. Loneliness and Risk of Dementia. J Gerontol Ser B 2018 Nov 16;Innovation in Aging, Volume 2(suppl_1):966-967. [doi: 10.1093/geroni/igy031.3581]
12. Simard J, Volicer L. Loneliness and Isolation in Long-term Care and the COVID-19 Pandemic. J Am Med Dir Assoc 2020 Jul;21(7):966-967 [FREE Full text] [doi: 10.1016/j.jamda.2020.05.006] [Medline: 32505516]
13. Chopik WJ. The Benefits of Social Technology Use Among Older Adults Are Mediated by Reduced Loneliness. Cyberpsychol Behav Soc Netw 2016 Sep;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
14. Neves BB, Franz RL, Munteanu C, Baecker R. Adoption and feasibility of a communication app to enhance social connectedness amongst frail institutionalized oldest old: an embedded case study. Inf Commun Soc 2017 Jul 13;21(11):1681-1699. [doi: 10.1080/1369118x.2017.1348534]
15. Lee C, Coughlin JF. Perspective: Older Adults' Adoption of Technology: An Integrated Approach to Identifying Determinants and Barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
16. Chin WW, Johnson N, Schwarz A. A Fast Form Approach to Measuring Technology Acceptance and Other Constructs. MIS Q 2008;32(4):687. [doi: 10.2307/25148867]
17. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]

18. Barbosa Neves B, Franz R, Judges R, Beermann C, Baecker R. Can Digital Technology Enhance Social Connectedness Among Older Adults? A Feasibility Study. J Appl Gerontol 2019 Jan;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]

Abbreviations

CNIL: National Commission for Information Technology and Civil Liberties
TOVID: Telephony Or Videophony for Isolated elderly

Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by C Miranda, I David; comments to author 15.07.20; revised version received 17.07.20; accepted 22.07.20; published 18.09.20.

Please cite as:
Sacco G, Lléonart S, Simon R, Noublanche F, Annweiler C, TOVID Study Group
Communication Technology Preferences of Hospitalized and Institutionalized Frail Older Adults During COVID-19 Confinement: Cross-Sectional Survey Study
JMIR Mhealth Uhealth 2020;8(9):e21845
URL: http://mhealth.jmir.org/2020/9/e21845/
doi:10.2196/21845
PMID:32896832

Original Paper

Association of Social Network Use With Increased Anxiety Related to the COVID-19 Pandemic in Anesthesiology, Intensive Care, and Emergency Medicine Teams: Cross-Sectional Web-Based Survey Study

Thomas Clavier[1], MD, PhD; Benjamin Popoff[1], MSc; Jean Selim[1], MD; Marion Beuzelin[2], MD; Melanie Roussel[3], MD; Vincent Compere[1], MD, PhD; Benoit Veber[1], MD, PhD; Emmanuel Besnier[1], MD, PhD

[1]Department of Anesthesiology and Critical Care, Rouen University Hospital, Rouen, France

[2]Department of Critical Care, Dieppe General Hospital, Dieppe, France

[3]Department of Emergency Medicine, Rouen University Hospital, Rouen, France

Corresponding Author:
Thomas Clavier, MD, PhD
Department of Anesthesiology and Critical Care
Rouen University Hospital
1 rue de Germont
Rouen, 76000
France
Phone: 33 685656293
Email: thomasclavier76@gmail.com

Abstract

Background: Critical care teams are on the front line of managing the COVID-19 pandemic, which is stressful for members of these teams.

Objective: Our objective was to assess whether the use of social networks is associated with increased anxiety related to the COVID-19 pandemic among members of critical care teams.

Methods: We distributed a web-based survey to physicians, residents, registered and auxiliary nurses, and nurse anesthetists providing critical care (anesthesiology, intensive care, or emergency medicine) in several French hospitals. The survey evaluated the respondents' use of social networks, their sources of information on COVID-19, and their levels of anxiety and information regarding COVID-19 on analog scales from 0 to 10.

Results: We included 641 respondents in the final analysis; 553 (86.3%) used social networks, spending a median time of 60 minutes (IQR 30-90) per day on these networks. COVID-19–related anxiety was higher in social network users than in health care workers who did not use these networks (median 6, IQR 5-8 vs median 5, IQR 3-7) in univariate (P=.02) and multivariate (P<.001) analyses, with an average anxiety increase of 10% in social network users. Anxiety was higher among health care workers using social networks to obtain information on COVID-19 than among those using other sources (median 6, IQR 5-8 vs median 6, IQR 4-7; P=.04). Social network users considered that they were less informed about COVID-19 than those who did not use social networks (median 8, IQR 7-9 vs median 7, IQR 6-8; P<.01).

Conclusions: Our results suggest that social networks contribute to increased anxiety in critical care teams. To protect their mental health, critical care professionals should consider limiting their use of these networks during the COVID-19 pandemic.

(JMIR Mhealth Uhealth 2020;8(9):e23153) doi:10.2196/23153

KEYWORDS

social network; nurse; physician; anxiety; emergency medicine, anesthesiology, critical care medicine; coronavirus disease 2019; mental health; COVID-19

Introduction

The emergence of SARS-CoV-2, which causes the disease COVID-19, at the end of 2019 caused a large global outbreak and represents a major public health issue. While 80% of patients appear to be mildly symptomatic or asymptomatic, about 20% develop viral pneumonia. Of these severe forms, 7% to 40% of patients progress to acute respiratory distress syndrome and may require admission to the intensive care unit (ICU) [1,2]. Although currently there are few data on the mortality of this infection, early reports suggest a high mortality rate of approximately 6% for all patients with COVID-19 and of up to 60% among patients admitted to the ICU [2,3].

Physicians and nurses providing critical care (eg, intensivists, anesthesiologists, and emergency physicians) are on the front line of management of the most severe forms of COVID-19. The viral load of SARS-CoV-2 detected in patients' respiratory tracts is positively linked to lung disease severity; therefore, patients with COVID-19 who are admitted to the emergency department or ICU are probably the most contagious [4]. Significant risk of transmission of SARS-CoV-2 from patients to health care workers has been described [5]. It is therefore obvious that given the risk of contamination and the high mortality rate of COVID-19, this pandemic is a great source of stress for health care workers [6].

Social networks (eg, Twitter and Facebook) enable users to find information by passively viewing a message or information thread without using traditional media, while instant messaging platforms (eg, WhatsApp) enable users to communicate directly with friends or colleagues. These apps are now commonly used by health care workers in many domains, such as teaching, promotion of scientific work, contact with patients, and discussion with colleagues [7-9]. Some health care workers based in areas that are currently severely affected by the COVID-19 pandemic (ie, northern Italy, eastern France, and New York City in the United States) use these tools to communicate and share their stressful experiences. Thus, since the beginning of the COVID-19 pandemic, many testimonials have emerged on social networks of issues such as lack of ICU beds, necessity to make difficult ethical decisions, and high numbers of deaths despite optimal care. Repeatedly reading such information can be a source of anxiety for health care workers who are already in contact with patients with COVID-19 or for those who have not yet been in contact with these patients. To preserve the mental health of these health care workers, who are essential to the functioning of the emergency organizations established in affected countries, it appears to be necessary to find strategies to limit the anxiety of health care workers. It is thus crucial to understand and analyze the sources of this anxiety.

The objective of this work was to assess whether the use of social networks and instant messaging apps to obtain and exchange information on the COVID-19 pandemic is associated with increased anxiety in critical care teams.

Methods

Population Selection

The Ethics and Evaluation Committee for Non-Interventional Research of Rouen University Hospital approved the study (No. E2020-12). We conducted a prospective study in France using a declarative survey. The link to an open, web-based Google Forms survey with 20 items on one webpage was sent by email to medical and paramedical teams in anesthesiology, intensive care, and emergency departments throughout France via professional or personal emailing lists (including lists belonging to department heads of the Rouen University Hospital). The survey was also distributed via professional WhatsApp discussion groups to which the authors belonged. Finally, two associative or academic societies (the Association of Young Anesthesiologists and Intensivists and the French Intensive Care Society) also relayed the questionnaire; the method of dissemination was left to the discretion of the community manager of each society. In practice, the questionnaire was disseminated to both societies via Twitter. All contacted health care workers were asked to forward the survey link to their colleagues. All participants received information about the survey objectives, which were recalled in the preface of the questionnaire. By voluntarily participating in the survey after receiving adequate information on its purpose, informed consent was implied. Although it was theoretically possible to identify individual participants, no efforts were made to do this, and no plausible harm to participating individuals could arise from the study. This survey was developed according to available guidelines for self-administered surveys [10]. Responses were entered on a single webpage with one Submit button that only allowed submissions via a unique link; thus, uninvited responses were extremely unlikely. The request was sent to 172 health care workers at our home institution; however, as we were unable to determine how many health care workers the request was forwarded to at other institutions, we do not know how many health care workers received the request to participate in the survey. The survey was conducted in accordance with the Checklist for Reporting Results of Internet E-surveys (CHERRIES) [11].

The participants included in the analysis were medical health care workers (physicians and residents) or paramedical health care workers (registered nurses, auxiliary nurses, and nurse anesthetists) who were in contact with patients in a critical care sector of a French hospital: anesthesiology, ICU, emergency department, mobile emergency, critical care unit, or mixed activity. Criteria for noninclusion were other health professions, health care workers who worked in other hospital departments or in a country other than France, and professionals who had no direct interaction with patients (eg, nursing managers).

Objectives

Our main objective was to compare the levels of anxiety related to the COVID-19 pandemic between health care workers who use social networks and those who do not use them.

The secondary objectives were to compare the levels of information related to the COVID-19 pandemic between health care workers who use social networks and those who do not use

them, to compare the levels of anxiety related to the COVID-19 pandemic between health care workers who use WhatsApp professionally and those who use WhatsApp but not professionally, and to compare the levels of anxiety related to the COVID-19 pandemic between respondents who use social networks as a source of information on the COVID-19 pandemic and those who use other information sources.

Survey Design

The survey was designed and written by TC and then reviewed, tested and validated by EB (assistant professor) and VC (full professor) before being sent. To avoid bias related to the evolution of anxiety as the pandemic progresses, the survey period lasted only one week. The survey was constructed in three parts. The Demographic Data section analyzed the participants' region and city of practice, age, gender, type of hospital (public or private, university), department, and profession. The Use of Social Networks section analyzed personal and professional use of WhatsApp, the social networks consulted at least once a week, and the average daily time spent on social networks. Finally, the Link Between Social Networks and the COVID-19 Pandemic section analyzed the sources of information participants used to learn about the pandemic, use of WhatsApp to discuss the pandemic with colleagues, participation in a WhatsApp group dedicated solely to discussions about COVID-19, direct interaction with a patient with COVID-19, work in a hospital that was already in contact with or was expecting to be in contact with patients with COVID-19, and subjective levels of information and anxiety about the pandemic (rated on a Likert scale from 0 to 10).

For the demographic analysis, the 14 regions of practice (13 French metropolitan regions and 1 overseas territory) were grouped into two areas according to the incidence rate of COVID-19 on March 19, 2020 (the day before the survey was released): low-density areas (regions with an incidence rate below the median incidence rate in France) and high-density areas (regions with an incidence rate above the median incidence rate in France; see Table 1).

The web-based survey can be accessed on the internet in French [12]. A version of the survey translated into English is available in Multimedia Appendix 1.

Table 1. Clustering of French regions in low- and high-density areas according to COVID-19 incidence.

French area and region	Cases per 100,000 inhabitants (March 19, 2020)
COVID-19 low-density areas	
Pays de la Loire	6.15
Outre-mer (Martinique, Guadeloupe, Guyane, Réunion, Mayotte)	7.80
Nouvelle-Aquitaine	8.68
Centre-Val de Loire	9.93
Occitanie	10.31
Normandie	10.44
Bretagne	10.75
COVID-19 high-density areas	
Hauts-de-France	13.27
Auvergne-Rhône-Alpes	15.76
Provence-Alpes-Côte d'Azur	20.59
Bourgogne-Franche-Comté	32.45
Île-de-France	38.24
Corse	50.19
Grand Est	56.04

Statistical Analyses

In univariate analyses, chi-square tests were performed for categorical variables and Wilcoxon tests were performed for continuous variables. A multivariable analysis by linear regression was performed to model the anxiety score and adjust for confounders, including age, region density, type of department, level of information on COVID-19, and whether the department was already providing (or was going to provide) care for patients with COVID-19. The Spearman rank correlation test was used to assess the association between two variables. The analyses were conducted bilaterally, taking a significance threshold of $P<.05$. Continuous variables were described as median (IQR) and categorical variables were described as absolute numbers and percent prevalence (%). All statistical analyses were performed using R version 3.5.1 (R Foundation for Statistical Computing).

Data Availability

The raw data supporting the conclusions of this manuscript can be made available on request by the authors to any qualified researcher.

Results

Demographic Characteristics

The responses were compiled from March 20 to 27, 2020. A total of 759 health care workers responded to the survey; 118 respondents did not meet our inclusion criteria and were excluded from the analysis (Figure 1). Among the 641 respondents analyzed, the median age was 33 years (IQR 29-41), the sex ratio was 0.79 (282 men, 359 women), and 170 (26.5%) worked in a COVID-19 high-density area. The respondents' main sources of information on the COVID-19 pandemic were discussion with colleagues, institutional information (from hospital management or academic societies), and scientific literature (Figure 2).

Figure 1. Flowchart of the selection of the study respondents.

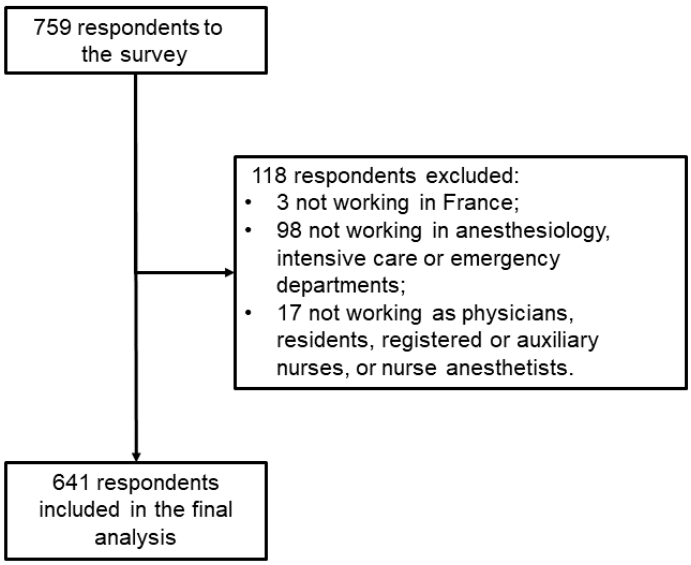
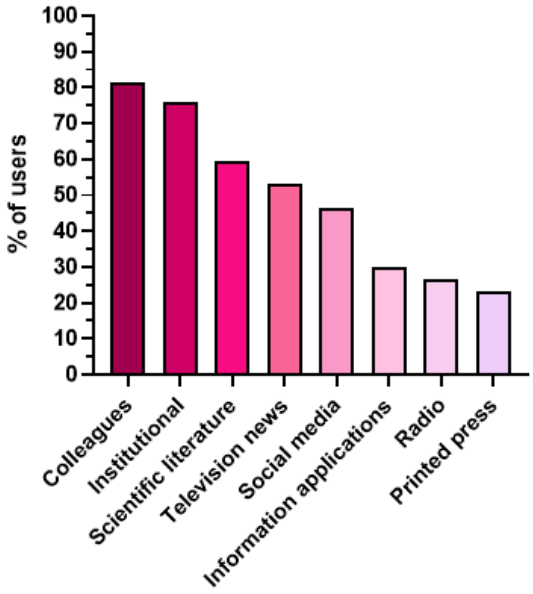


Figure 2. Percentages of respondents who used different media to obtain information on the COVID-19 pandemic.



Use of Social Networks and Anxiety

Among the 641 respondents included in the final analysis, 553 (86.3%) were social network users, and they spent a median time of 60 minutes (IQR 30-90) per day on these networks. The social networks used by the respondents are detailed in Figure 3. In the univariate analysis, respondents who used social networks worked more in COVID-19 low-density areas and intensive care units, and they were more likely to work in a hospital that was managing patients with COVID-19 (Table 2). They also reported a higher level of anxiety regarding COVID-19 and felt less informed than respondents who did not use social networks (Table 2). In multivariate analysis adjusted for age, density of region, type of department, level of information on COVID-19, and whether the department provided care for patients with COVID-19, anxiety was significantly associated with social network use (P<.001), with an average increase of 1.0 anxiety point (corresponding to a 10% increase) in social network users.

Among social network users, the level of anxiety was higher among health care workers who used social networks to obtain information on the COVID-19 pandemic than among those who used other sources of information (median 6, IQR 5-8 vs median 6, IQR 4-7; *P*=.04). There was no correlation between the time spent on social networks and the level of anxiety (*r*=0.08, 95% CI 0.00 to 0.17; *P*=.05) or the level of information (*r*=0.02, 95% CI –0.06 to 0.11; *P*=.57). There was no difference in anxiety levels between health care workers who used WhatsApp to discuss COVID-19 with other professionals (median 6, IQR 5-7) and those who did not use it (median 6, IQR 4-8; *P*=.32). There was no correlation between the information level concerning the COVID-19 pandemic and the level of anxiety (*r*=–0.05, 95% CI –0.13 to 0.03; *P*=.17).

Figure 3. Percentages of users who consulted different social networks at least once per week (n=553).

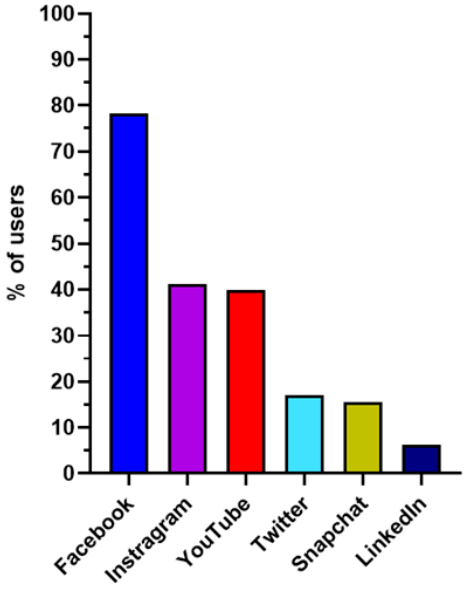


Table 2. Comparisons of the characteristics and use of social networks by health care workers (N=641). Percentages are expressed in relation to the number of respondents to each question.

Characteristic	Health care workers not using social networks	Health care workers using social networks	*P* value
n (%)	88 (13.7)	553 (86.3)	N/A[a]
Age (years), median (IQR)	46.0 (35.0-54.3)	32.0 (28.0-38.0)	<.001
Sex ratio (male/female)	0.87	0.96	.68
COVID-19 low-density area, n (%)	73 (83.0)	398 (72.0)	.04
Profession, n (%)			.41
Physician or resident	62 (70.5)	361 (65.3)	
Nurse (registered or auxiliary) or nurse anesthetist	26 (29.5)	192 (34.7)	
Department, n (%)			.02
Anesthesiology	42 (47.7)	181 (32.7)	
Intensive care unit	23 (26.1)	223 (40.3)	
Emergency	6 (6.8)	55 (9.9)	
Mobile emergency	9 (10.2)	34 (6.1)	
Mixed	8 (9.1)	60 (10.8)	
Hospital, n (%)			.54
University	52 (59.1)	325 (58.8)	
Public	30 (34.1)	180 (32.5)	
Private	6 (6.8)	35 (6.3)	
Working in a hospital providing care to patients with COVID-19, n (%)	65 (73.9)	474 (85.7)	<.01
Direct interaction with patients with COVID-19, n (%)	33 (37.5)	240 (43.4)	.36
Use of WhatsApp, n (%)	81 (92.0)	517 (93.5)	.78
Use of WhatsApp to discuss COVID-19 with critical care professionals, n (%)	61 (73.5)	379 (71)	.73
Use of a WhatsApp group specifically dedicated to COVID-19, n (%)	41 (50.0)	213 (40.0)	.11
Anxiety level concerning the COVID-19 pandemic (from 0 to 10), median (IQR)	5 (3-7)	6 (5-8)	.02
Information level concerning the COVID-19 pandemic (from 0 to 10), median (IQR)	8 (7-9)	7 (6-8)	<.01

[a]N/A: not applicable.

Discussion

Principal Results

We have shown that the use of social networks is independently associated with a higher level of anxiety among health care workers in critical care sectors. To our knowledge, we describe a link between the use of social networks and anxiety in medical teams in a pandemic context for the first time.

Comparison With Prior Work

Several studies have shown that the COVID-19 pandemic is causing increased anxiety both in infected patients and in professionals who are in contact with these patients. Thus, rates of moderate or major anxiety ranging from 9% to 46% have been reported in health workers involved in the management of patients with of COVID-19, depending on the study [13,14]. It has been shown that in this population, age <35 years, female sex, and working in contact with patients with COVID-19 for >3 hours per day are associated with the development of an anxiety disorder due to COVID-19 [15,16]. These characteristics correspond to our population; therefore, we can assume that the population we studied is at high risk for pandemic-induced anxiety disorders.

In our study, the rate of social network users was high but consistent with data from previous studies that reported rates ranging from 50% to 88% among health care professionals [17-19]. The rate of WhatsApp users was also similar to that in a recent report, in which it was found that 98% of primary health care professionals used WhatsApp [17]. Social networks are known to be involved in the dissemination of “fake news” and rumors [20]. It has also been shown that on Facebook and YouTube, during a previous pandemic situation (Zika virus), misleading posts and videos were far more popular than those containing accurate and relevant public health information about

the disease [21,22]. It is therefore very likely that many misinformation or conspiracy theories about the COVID-19 pandemic are circulating on social networks; for this reason, the World Health Organization recently warned against "trolls and conspiracy theories" about COVID-19 [23]. Thus, it is interesting to note that in our work, respondents who used social networks felt less informed than their colleagues who did not use them, and respondents who used social networks as a source of information on the COVID-19 pandemic showed a higher level of anxiety. Our data suggest that social networks are inefficient in providing quality information and that the information provided is stressful. However, even if we did not find a correlation between the respondents' level of information concerning COVID-19 and level of anxiety in our study, it is also possible that the use of social networks to obtain information is linked to a lack of information, which itself can potentially cause anxiety. Because the design of our study did not allow us to establish a causal link, it would have been interesting to conduct interviews with a few caregivers using social networks to determine the causal link between these parameters.

The association between anxiety or depression and social network use has already been described in patients and in the general population [24,25]. However, this is the first study that specifically highlights this association in members of critical care teams. It is possible that false information (or the highly anxiety-inducing formulation of true information) may increase the anxiety experienced by health care workers; this would explain our results, at least in part. Despite the fact that time spent on social networks and anxiety score were not correlated in our study (P=.05), there are reports in the literature that suggest a dose-dependent relationship between time spent on social networks and anxiety [25,26]. It is known that work-related anxiety is associated with many complications in health care workers (particularly in critical care sectors), including accidents, medical errors, burnout, and secondary traumatic stress [27,28]. The COVID-19 pandemic is clearly associated with high levels of stress, sleeping disorders, and anxiety among health care professionals who interact with these patients [6,29]. It is therefore essential for these professionals to manage their mental health and to limit their sources of stress. Our results suggest that advising health care professionals to limit (or even temporarily stop) their consultation of social networks could be a way to limit this professional stress related to COVID-19.

Limitations

Despite our interesting results, our work has several major limitations. First, because our study was observational, it is not possible to establish a causal link between the use of social networks and anxiety. This is a cross-sectional study; therefore, we did not analyze the evolution of anxiety over time. It was shown in a longitudinal study conducted over a 4-week period during the COVID-19 pandemic that the anxiety of the general population did not vary significantly [30]. However, it is questionable whether this conclusion applies to health care workers. Second, our study focused on social networks but was itself partly disseminated on Twitter, which causes selection bias and explains the high rates of social network users and young health care workers in our work. Moreover, people who responded to the Google survey may be more familiar with the use of social networks than other health care workers. However, the use of conventional email diffusion allowed us to constitute a control group with a significant number of respondents and a rate of social network users that is comparable to those described among health professionals [17-19]. Third, we estimated anxiety with a single question on a Likert scale without considering multiple psychological components (depression, anxiety, insomnia, and distress) and without using a specific neuropsychological test as previously described [6]. The study was conducted in the context of major work overloads for critical care teams in some French regions. It is therefore very likely that some health care workers did not have time to answer the questionnaire. To be able to obtain responses from high-density sectors, we therefore deliberately chose to limit ourselves to a small number of questions and to greatly simplify our assessment of anxiety. However, it has been shown that evaluation by a computerized anxiety visual analog scale is reliable and is correlated with more complex neuropsychological tests [31]. We therefore considered that our assessment of anxiety was interpretable. Fourth, the questionnaire was written, proofread, and sent out in a few days in the middle of a peak of the epidemic, and we did not have time to have it validated by sending it to experts in survey design, anxiety, or pandemic infectious disease. This may limit the relevance of some of the items analyzed (eg, the item "discussion with colleagues" does not differentiate between whether these conversations involve the use of social networks or not). Fifth, by taking data (age, sex, profession) for individual respondents and cross-referencing the information, if one has the list of names and roles of the people working in all the hospitals, there is a theoretical possibility that some participants could be identified. The anonymization of the respondents is therefore incomplete; however, no efforts were made to identify individual participants in this study. Sixth, we analyzed anxiety in health care workers; however, it is possible that because they receive superior information about the reality of the pandemic, the anxiety of health care workers may be lower than that of the general population. Recent studies have found COVID-19 pandemic anxiety syndrome rates of 29% to 45% in the general population, while a meta-analysis determined that this anxiety was only detected in 23% of health care workers [32-34]. Finally, we only collected the potentially harmful effects of social networks; however, it is also possible that, as recently suggested, use of these networks has beneficial effects that we did not record in this work, such as preservation of contact with family and friends during confinement or relaxation while watching videos [35]. It has already been shown that social networks can also be used to disseminate valid information from public health organizations [36]. In future, it would be interesting to evaluate the potential beneficial effects of these networks during the COVID-19 pandemic.

Conclusion

Although we did not establish a direct causal link, in this paper, we have shown that the use of social networks is independently associated with increased anxiety among health care workers involved in the management of patients with severe COVID-19.

To protect their mental health, critical care professionals who are already under intense stress because of the current pandemic may want to limit their use of these networks during the coming months.

Acknowledgments

The authors are grateful to Nikki Sabourin-Gibbs (Rouen University Hospital) for her help in editing the manuscript. Funding support was provided solely from departmental sources.

Authors' Contributions

TC was involved in the study conception and design, in acquisition of the data, in the analysis and interpretation of the data, and in drafting the manuscript. BP was involved in the statistical analysis, in the analysis and interpretation of the data, and in drafting the manuscript. JS was involved in the acquisition and interpretation of the data and in drafting the manuscript. MB and MR were involved in the study conception and design, in acquisition of the data, and in drafting the manuscript. BV and VC were involved in the study conception and design, in interpretation of the data, and in revising the manuscript. EB was involved in the study conception and design, in the analysis and interpretation of the data, and in revising the manuscript. All authors contributed to the manuscript revision and have read and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English translation of the survey.
[DOCX File , 17 KB - mhealth_v8i9e23153_app1.docx]

References

1. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med 2020 Jul 01;180(7):934-943 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0994] [Medline: 32167524]
2. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481 [FREE Full text] [doi: 10.1016/S2213-2600(20)30079-5] [Medline: 32105632]
3. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. Lancet Infect Dis 2020 Jul;20(7):773. [doi: 10.1016/s1473-3099(20)30195-x]
4. Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. Sci China Life Sci 2020 Mar;63(3):364-374 [FREE Full text] [doi: 10.1007/s11427-020-1643-8] [Medline: 32048163]
5. Ng K, Poon BH, Kiat Puar TH, Shan Quah JL, Loh WJ, Wong YJ, et al. COVID-19 and the Risk to Health Care Workers: A Case Report. Ann Intern Med 2020 Jun 02;172(11):766-767. [doi: 10.7326/L20-0175] [Medline: 32176257]
6. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
7. Clavier T, Ramen J, Dureuil B, Veber B, Hanouz J, Dupont H, et al. Use of the Smartphone App WhatsApp as an E-Learning Method for Medical Residents: Multicenter Controlled Randomized Trial. JMIR mHealth uHealth 2019 Apr 09;7(4):e12825 [FREE Full text] [doi: 10.2196/12825] [Medline: 30964435]
8. Demailly Z, Brulard G, Selim J, Compère V, Besnier E, Clavier T. Gender differences in professional social media use among anaesthesia researchers. Br J Anaesth 2020 Mar;124(3):e178-e184. [doi: 10.1016/j.bja.2019.12.030] [Medline: 31987471]
9. Rolls K, Hansen M, Jackson D, Elliott D. How Health Care Professionals Use Social Media to Create Virtual Communities: An Integrative Review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
10. Burns KEA, Duffett M, Kho ME, Meade MO, Adhikari NKJ, Sinuff T, ACCADEMY Group. A guide for the design and conduct of self-administered surveys of clinicians. CMAJ 2008 Jul 29;179(3):245-252. [doi: 10.1503/cmaj.080372] [Medline: 18663204]
11. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
12. Réseaux sociaux et Covid-19. URL: https://docs.google.com/forms/d/e/1FAIpQLSdY_JnGQlxxzm0LH4hPzKSM5BO8B10kgoI4F5xs8U4C7uPceQ/viewform?usp=sf_link [accessed 2020-09-18]

13. Chew NWS, Lee GKH, Tan BYQ, Jing M, Goh Y, Ngiam NJH, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Aug;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]
14. Que J, Shi L, Deng J, Liu J, Zhang L, Wu S, et al. Psychological impact of the COVID-19 pandemic on healthcare workers: a cross-sectional study in China. Gen Psychiatr 2020;33(3):e100259 [FREE Full text] [doi: 10.1136/gpsych-2020-100259] [Medline: 32596640]
15. Lin Y, Hu Z, Alias H, Wong LP. Knowledge, Attitudes, Impact, and Anxiety Regarding COVID-19 Infection Among the Public in China. Front Public Health 2020;8:236 [FREE Full text] [doi: 10.3389/fpubh.2020.00236] [Medline: 32574305]
16. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
17. Marin-Gomez FX, Garcia Cuyas F, Reig-Bolano R, Mendioroz J, Roura-Poch P, Pico-Nicolau M, et al. Social Networking App Use Among Primary Health Care Professionals: Web-Based Cross-Sectional Survey. JMIR mHealth uHealth 2018 Dec 21;6(12):e11147 [FREE Full text] [doi: 10.2196/11147] [Medline: 30578175]
18. Lefebvre C, McKinney K, Glass C, Cline D, Franasiak R, Husain I, et al. Social Media Usage Among Nurses: Perceptions and Practices. J Nurs Adm 2020 Mar;50(3):135-141. [doi: 10.1097/NNA.0000000000000857] [Medline: 32049701]
19. Surani Z, Hirani R, Elias A, Quisenberry L, Varon J, Surani S, et al. Social media usage among health care providers. BMC Res Notes 2017 Nov 29;10(1):654 [FREE Full text] [doi: 10.1186/s13104-017-2993-y] [Medline: 29187244]
20. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science 2019 Jan 25;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
21. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
22. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health 2018 Sep;112(6):320-328 [FREE Full text] [doi: 10.1080/20477724.2018.1507784] [Medline: 30156974]
23. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3) [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
24. Shensa A, Sidani JE, Dew MA, Escobar-Viera CG, Primack BA. Social Media Use and Depression and Anxiety Symptoms: A Cluster Analysis. Am J Health Behav 2018 Mar 01;42(2):116-128 [FREE Full text] [doi: 10.5993/AJHB.42.2.11] [Medline: 29458520]
25. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association Between Social Media Use and Depression Among U.S. Young Adults. Depress Anxiety 2016 Apr;33(4):323-331 [FREE Full text] [doi: 10.1002/da.22466] [Medline: 26783723]
26. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017 Jan 01;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]
27. Weaver MD, Vetter C, Rajaratnam SMW, O'Brien CS, Qadri S, Benca RM, et al. Sleep disorders, depression and anxiety are associated with adverse safety outcomes in healthcare workers: A prospective cohort study. J Sleep Res 2018 Dec;27(6):e12722 [FREE Full text] [doi: 10.1111/jsr.12722] [Medline: 30069960]
28. van Mol MMC, Kompanje EJO, Benoit DD, Bakker J, Nijkamp MD. The Prevalence of Compassion Fatigue and Burnout among Healthcare Professionals in Intensive Care Units: A Systematic Review. PLoS One 2015;10(8):e0136955 [FREE Full text] [doi: 10.1371/journal.pone.0136955] [Medline: 26322644]
29. Xiao H, Zhang Y, Kong D, Li S, Yang N. The Effects of Social Support on Sleep Quality of Medical Staff Treating Patients with Coronavirus Disease 2019 (COVID-19) in January and February 2020 in China. Med Sci Monit 2020 Mar 05;26:e923549 [FREE Full text] [doi: 10.12659/MSM.923549] [Medline: 32132521]
30. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
31. Abend R, Dan O, Maoz K, Raz S, Bar-Haim Y. Reliability, validity and sensitivity of a computerized visual analog scale measuring state anxiety. J Behav Ther Exp Psychiatry 2014 Dec;45(4):447-453. [doi: 10.1016/j.jbtep.2014.06.004] [Medline: 24978117]
32. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]
33. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5) [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]

34. Bäuerle A, Teufel M, Musche V, Weismüller B, Kohler H, Hetkamp M, et al. Increased generalized anxiety, depression and distress during the COVID-19 pandemic: a cross-sectional study in Germany. J Public Health (Oxf) 2020 Jul 13 [FREE Full text] [doi: 10.1093/pubmed/fdaa106] [Medline: 32657323]
35. Wiederhold BK. Using Social Media to Our Advantage: Alleviating Anxiety During a Pandemic. Cyberpsychol Behav Soc Netw 2020 Apr;23(4):197-198. [doi: 10.1089/cyber.2020.29180.bkw] [Medline: 32207630]
36. Daughton AR, Paul MJ. Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus. J Med Internet Res 2019 May 13;21(5):e13090 [FREE Full text] [doi: 10.2196/13090] [Medline: 31094347]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-surveys
ICU: intensive care unit

Edited by L Buis, G Eysenbach; submitted 03.08.20; peer-reviewed by R Ho, R Kadri; comments to author 01.09.20; revised version received 03.09.20; accepted 04.09.20; published 24.09.20.

Please cite as:
Clavier T, Popoff B, Selim J, Beuzelin M, Roussel M, Compere V, Veber B, Besnier E
Association of Social Network Use With Increased Anxiety Related to the COVID-19 Pandemic in Anesthesiology, Intensive Care, and Emergency Medicine Teams: Cross-Sectional Web-Based Survey Study
JMIR Mhealth Uhealth 2020;8(9):e23153
URL: http://mhealth.jmir.org/2020/9/e23153/
doi:10.2196/23153
PMID:32924946

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org