JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 7 (2019), Issue 10 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Original Papers

Reviews

Corrigenda and Addenda

Original Paper

The Fever Coach Mobile App for Participatory Influenza Surveillance in Children: Usability Study

Myeongchan Kim[1*], MD; Sehyo Yune[1*], MD, MPH, MBA; Seyun Chang[1], BS; Yuseob Jung[2,3], BS; Soon Ok Sa[2,3], MS; Hyun Wook Han[2,3,4], MD, PhD

[1]Mobile Doctor, Co, Ltd, Seoul, Republic of Korea

[2]Institute of Basic Medical Sciences, Graduate School of Medicine, CHA University, Seongnam-si, Gyeonggi-do, Republic of Korea

[3]Department of Biomedical Informatics, Graduate School of Medicine, CHA University, Seongnam-si, Gyeonggi-do, Republic of Korea

[4]Healthcare Bigdata Center, Bundang CHA General Hospital, Seongnam-si, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyun Wook Han, MD, PhD
Department of Biomedical Informatics
Graduate School of Medicine, CHA University
Pangyo-ro 335, Bundang-gu
Seongnam-si, Gyeonggi-do, 13488
Republic of Korea
Phone: 82 31 881 7109
Email: stepano7@gmail.com

Abstract

Background: Effective surveillance of influenza requires a broad network of health care providers actively reporting cases of influenza-like illnesses and positive laboratory results. Not only is this traditional surveillance system costly to establish and maintain but there is also a time lag between a change in influenza activity and its detection. A new surveillance system that is both reliable and timely will help public health officials to effectively control an epidemic and mitigate the burden of the disease.

Objective: This study aimed to evaluate the use of parent-reported data of febrile illnesses in children submitted through the Fever Coach app in real-time surveillance of influenza activities.

Methods: Fever Coach is a mobile app designed to help parents and caregivers manage fever in young children, currently mainly serviced in South Korea. The app analyzes data entered by a caregiver and provides tailored information for care of the child based on the child's age, sex, body weight, body temperature, and accompanying symptoms. Using the data submitted to the app during the 2016-2017 influenza season, we built a regression model that monitors influenza incidence for the 2017-2018 season and validated the model by comparing the predictions with the public influenza surveillance data from the Korea Centers for Disease Control and Prevention (KCDC).

Results: During the 2-year study period, 70,203 diagnosis data, including 7702 influenza reports, were submitted. There was a significant correlation between the influenza activity predicted by Fever Coach and that reported by KCDC (Spearman ρ=0.878; P<.001). Using this model, the influenza epidemic in the 2017-2018 season was detected 10 days before the epidemic alert announced by KCDC.

Conclusions: The Fever Coach app successfully collected data from 7.73% (207,699/2,686,580) of the target population by providing care instruction for febrile children. These data were used to develop a model that accurately estimated influenza activity measured by the central government agency using reports from sentinel facilities in the national surveillance network.

(JMIR Mhealth Uhealth 2019;7(10):e14276) doi:10.2196/14276

KEYWORDS

data collection; detecting epidemics; mobile app; health care app; influenza epidemics; influenza in children

Introduction

Seasonal influenza affects 5% to 15% of the world population and is estimated to be accountable for 3 to 5 million cases of severe illness and up to 500,000 deaths annually [1]. Children younger than 5 years, adults aged 65 years or older, and pregnant women are at high risk of developing serious influenza-related complications [2]. In 2008, it was estimated that 90 million new cases, 1 to 2 million cases of severe illness, and 28,000 to 111,500 deaths that were attributable to influenza infection occurred in this population [3].

Monitoring influenza activities is essential to understand the epidemiology, implement appropriate prevention strategies, and adequately allocate public health resources to mitigate the burden of epidemics. Currently, most developed countries have surveillance systems based on the networks of sentinel facilities and primary care practitioners reporting the weekly number of patients visited with influenza-like illness (ILI) or acute respiratory infection. Many countries, including the United States and South Korea, also collect data of virologic information, hospitalization, and mortality. At a global level, the World Health Organization Global Influenza Surveillance and Response System, comprising 143 National Influenza Centers in 111 countries, has provided virologic information used to select strains for influenza vaccine formulations [4,5].

Although the information provided by influenza surveillance is invaluable, a traditional surveillance system requires substantial amount of resources, including trained staff, laboratory capacities, and information infrastructure. Establishing a sustainable surveillance system in countries with limited resources depends on financial and technical support from external sources as well as active involvement of the national health authority [6]. In addition, although timeliness is a crucial element of infection surveillance, the traditional system is associated with time lags as it takes several days from event to reports, from reports to analysis, and from analysis to dissemination [7].

To overcome the challenges of traditional surveillance systems, novel approaches to influenza surveillance using the internet and mobile technologies have been developed. The use of big data drawn from Web search queries, social media, and electronic health records has attracted much attention from public health researchers and policy makers. Among these, Google Flu Trends (GFT) [8] was one of the key drivers of the hype around big data in public health. Despite its constant overestimations of flu prevalence that led to the termination of the service, GFT shed light on the use of internet big data in disease surveillance and nowcasting [9,10]. A number of researchers have also evaluated search queries and social media as influenza surveillance tools [11-15]. However, the big data approach faces serious challenges including *big data hubris*, privacy concerns, and a lack of accurate algorithmic models [9,10].

Another recently emerged internet-based surveillance method is participatory surveillance [16]. Participatory disease surveillance systems ask the target population to submit disease-related data through various forms of survey tools. Although it is a potentially valuable data source, participatory surveillance has low specificity because many such systems collect subjective symptoms of illness without information on clinical diagnosis [16,17]. Furthermore, as these systems completely rely on volunteers in communities, participation is low and difficult to maintain over time [18].

In this paper, we describe a new approach that collects participatory data, including symptoms and clinical diagnosis, of febrile diseases and encourages participation by providing fever care instructions. In addition to analyzing individual data and delivering customized advice, the vast amount of aggregate information could generate a population-level data. We examined if such data could be used for influenza surveillance while addressing the aforementioned challenges of participatory disease surveillance systems.

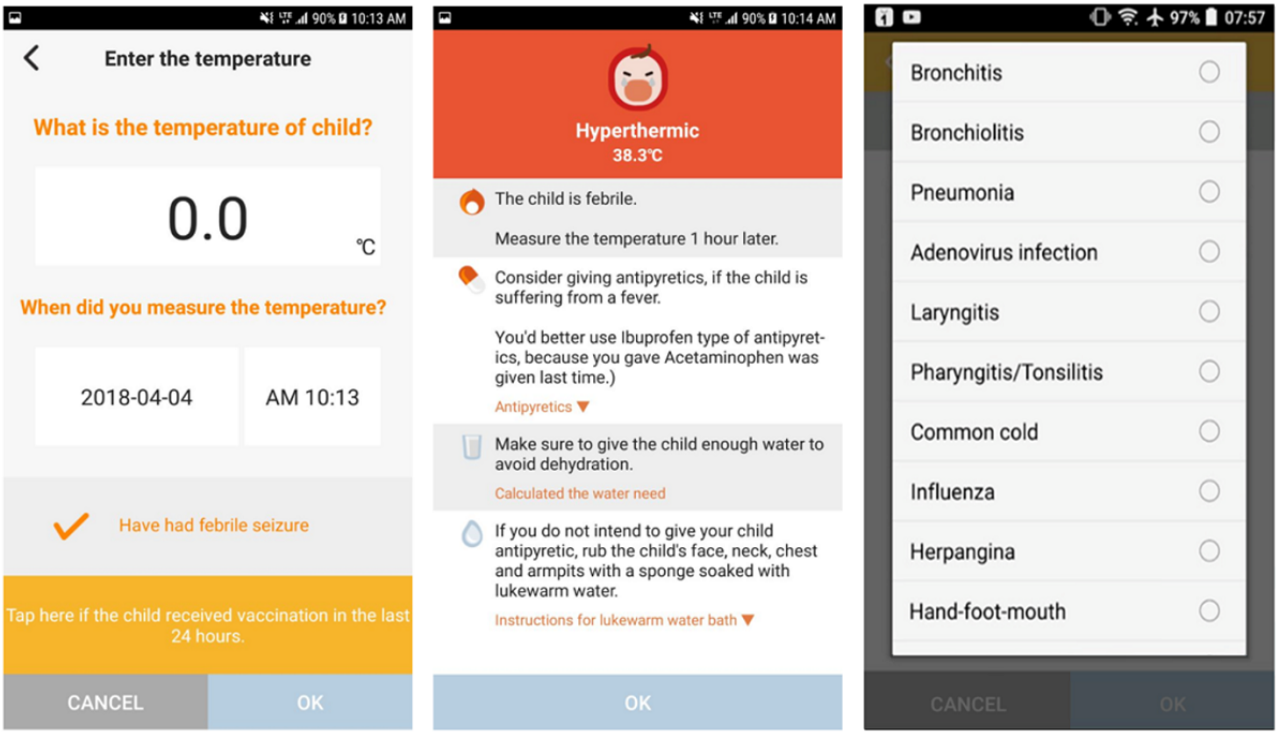
Methods

Exploratory Data Analysis

Participatory Data: Fever Coach

All self-reported data were obtained from Fever Coach. Fever Coach (in Korean) is a mobile app created to provide actionable, tailored information for fever management in children younger than 5 years. Upon registration for the service, a user is asked to enter the child's date of birth, body weight, sex, and history of febrile seizure. A single user can manage multiple accounts if they have more than 1 child. Figure 1 shows screenshots of the app. After registration, the user may enter various fever-related information of the child including body temperature, other symptoms, dose and time of antipyretic administration, vaccination and antibiotic history, and physician-made diagnosis if the child was seen by a physician for the management of current illness. On the basis of these inputs, the app provides instructions for fever control and advises on the timing of seeking medical attention. If a physician-made diagnosis is submitted, detailed information about the disease is also provided. Users can also choose to not submit any data and to only read fever care instructions. The service was designed and reviewed by 2 board-certified family physicians and 1 board-certified pediatrician. Fever Coach was launched in July 2015 in South Korea. The app is currently available for free on Google Play and Apple App Store in South Korea, Japan, and China. As of June 2018, there have been 400,000 downloads, 99% of which were in South Korea.

Figure 1. Screenshots of Fever Coach app. (a) body temperature page, (b) fever management page, (c) diagnosis page. The service is not provided in English and is translated from Korean for these screenshots. (options list: bronchitis, bronchiolitis, pneumonia, adenovirus infection, laryngitis, pharyngitis/tonsillitis, common cold, influenza, canker sore, herpangina, hand-foot-mouth, enterocolitis, meningitis, otitis media, urinary tract infection, mumps, scarlet fever, chicken pox, measles, roseola, not diagnosed, and others).



The study data were collected retrospectively only from users who agreed to the use of deidentified information for research purposes from September 2016 to August 2018. The data submitted between September 2016 and August 2017 (development period, 52 weeks) were used to develop a prediction model, and the data submitted from September 2017 to August 2018 (validation period, 52 weeks) were used to validate the model.

All user-reported data were stored in Fever Coach databases in real time, and metadata (number of submissions and location information) on each day were processed and saved on a daily basis. For comparison with the available reference population, only the data from children aged 1 year (365 days) to 6 years (2554 days) submitted from South Korea were included for this study. To filter out falsely submitted data, we only used body weight values from 5 percentile of newborns (2.5 kg) to 95 percentile of 18 year olds (79.8 kg) based on the Korean child growth curve in 2017 [19]. As the age of a child changed over time, we calculated the mean of the age during each period if data were submitted multiple times. For body weight, we only considered the last submitted data for each child. These data were used to understand user demographics, clinical characteristics, and fever-related behavior. The study protocol was approved by the institutional review board (IRB) of CHA university (IRB number 1044308-201804-HR-022-03).

The total app user activity was measured by the number of body temperature submissions to the app in each week. As Fever Coach is typically used only when a child is ill, the authors decided that it was more appropriate to measure user activity by the number of data submissions than by daily or weekly active users. The proportion of influenza diagnoses to the total number of any diagnoses submitted to the app was used as the indicator of influenza activity.

Reference Data: Korea Centers for Disease Control and Prevention Surveillance System

The KCDC ILI surveillance system provides weekly updates on influenza surveillance data as the proportion of the ILI-related visits to the total outpatient visits [20]. An ILI is defined as a fever equal to or greater than 38.0°C (100.4°F) accompanied by cough or sore throat. The surveillance data are collected from the Korea Influenza Surveillance Scheme that has 36 sentinel primary care facilities. KCDC also reports the proportion of the ILI-related visits of children aged 1 year through 6 years to the total outpatient visits of the same age range in a week, as well as in the entire age range.

Prediction of Influenza Activities

Prediction Model

A linear regression model was developed to predict influenza activity in children from the Fever Coach data:

$$P(t)=\beta_0 + \beta_1 F(t) + \varepsilon$$

$P(t)$ is the proportion of the ILI-related visits of children aged 1 year through 6 years to the total outpatient visits of the same-aged population reported by KCDC, $F(t)$ and is the proportion of influenza diagnosis reports to all diagnosis reports through the Fever Coach app. Both $P(t)$ and $F(t)$ include data accrued in 7 days that precedes date t Intercept β_0 and slope β_1

were obtained from regression, and ε is an error term. We also developed another linear regression model on $F(t)$ to predict the proportion of ILI visits in all age ranges to detect the influenza epidemic alert, which is based on the ILI visits in the total population. The threshold of KCDC for an influenza epidemic alert in the 2017 flu season was set at 6.6 ILI visits per 1000 visits, determined as the average value of ILI plus 2 SD during the nonflu seasons in the past 3 years (2014-2016).

Statistical Analysis

To compare user characteristics and behavior between the development period and the validation period, the Pearson chi-square test was used for categorical values, the Student *t* test was used for continuous values with normal distribution, and the Mann-Whitney U test was used for continuous values with non-normal distribution. The Shapiro-Wilk test was used for determining the normality of data distribution.

The accuracy of the prediction model was evaluated using the Spearman's rank correlation coefficient (Spearman ρ) and root mean square error (RMSE) for comparison between $P(t)$ and their predicted values. All statistical analyses were performed using *Scikit-learn 0.19.1* [21].

Results

Exploratory Data Analysis

During the entire study period, 10,002,512 fever-related health data were collected. The number of total records decreased by 18.5%, whereas the number of children whose data were submitted decreased by 44.2% from the development period to the validation period (Table 1).

Table 1. User characteristics and number of datapoints submitted during the development and validation periods.

Unit, category	Statistical items	Development period (9/1/2016-8/31/2017)	Validation period (9/1/2017-8/31/2018)
Datapoint	Total records	5,322,827	4,679,685
Body temperature	Number of records (%)	4,228,606 (79.44)	3,446,488 (73.65)
	Median (IQR[a])	38.0 (37.5-38.6)	38.0 (37.5-38.6)[b]
Antipyretics	Number of records (%)	856,210 (16.08)	1,096,443 (23.43)
Other symptoms	Number of records (%)	125,338 (2.35)	62,024 (1.33)
Antibiotics	Number of records (%)	38,122 (0.71)	27,841 (0.59)
Vaccination	Number of records (%)	24,243 (4.55)	10,828 (2.31)
Diagnosis	Total number	41,578	26,848
	Influenza, n (%)	2,965 (7.13)	5,800 (21.60)[b]
	Noninfluenza, n (%)	38,613 (92.97)	21,048 (78.37)[b]
Child	Total number	209,270	116,804
Number of data submission (per child)	Median (IQR)	12 (4-32)	21 (7-52)[b]
Age (years)	Median (IQR)	2.0 (1.4-3.1)	2.2 (1.5-3.3)[b]
Female sex, %	Ratio	49.22	49.26[c]
Weight (kg)	Median (IQR)	12.0 (10.7-14.6)	13.3 (11.5-16.0)[b]
Diagnosis submitted	Total number	27,736	18,128
	Influenza, n (%)	2455 (8.85)	5324 (29.37)
	Noninfluenza, n (%)	25,281 (91.15)	12,804 (70.63)

[a]IQR: interquartile range.

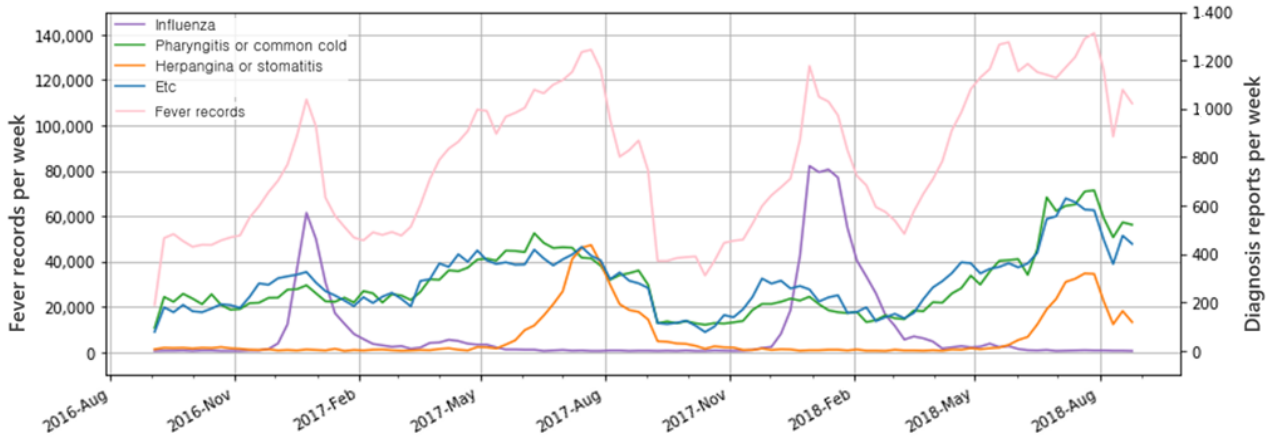
[b]Data showed significant difference between the 2 periods with $P<.001$. History of febrile seizure was not collected during the development period.

[c]$P=.807$.

During the development period, 149,329 people used the service at least once, and 115,674 (77.5%) of them used it for more than 2 separate days during the 1-year development period. From the 149,329 users, fever-related information was recorded from 209,270 children, and 49.2% (103,083) of them were female. This was 7.7% of the total population in that age group in South Korea [22]. They submitted 41,578 diagnoses, including 2965 influenza reports. The minimum and maximum values of $F(t)$ were 0 and 0.495, respectively, and the range of $P(t)$ was from 0.0037 to 0.0862 in the same period. The influenza activity during the study period is visualized geographically on a web page titled *Fever Coach Flu Visualization* [23]. The video record of this visualization can be found in Multimedia Appendix 1.

During the validation period, fever-related information of 116,804 children was recorded, and 26,848 diagnoses, including 5800 influenza reports, were collected. $F(t)$ and $P(t)$ ranged between 0 to 0.495 and 0.0023 to 0.0928, respectively. We classified all diagnosis records into 3 categories: influenza, pharyngitis/common cold, and herpangina/stomatitis. The change of diagnosis and app user activity over time is shown in Figure 2.

Figure 2. Change of diagnosis submissions and app user activity during the study period. Y-axis shows the number of submissions in each week. Fever records (left y-axis) was measured by the number of body temperature submissions to the app and is shown as a pink line. Number of each diagnosis categories (right y-axis) is shown with different colors. Pharyngitis or common cold (green line) shows the sum of pharyngitis ortonsillitis and common cold, and herpangina or stomatitis (orange line) shows the sum of canker sore and herpangina.



Prediction of Influenza Activities

The prediction model derived from the linear regression between ILI visits [$P(t)$] for children aged 1 year through 6 years and influenza diagnosis proportion submitted to Fever Coach [$F(t)$] came out with its slope (β_1) as 0.1316, intercept (β_0) as 0.0063, and R^2 as 0.930 (Figure 3). $P(t)$ is the reference value from KCDC. Spearman ρ was 0.895 (P<.001) and RMSE was 0.0083 between $P(t)$ and predicted value in the validation period (the maximum value of $P(t)$ was 0.0928 and minimum value was 0.0023). Figure 4 shows a graphical plot of $P(t)$ and predicted values over the study period as well as the error of the prediction model. From the all-age model that was developed between $F(t)$ and proportion of ILI visits in all age ranges, the was 0.924 (P<.001) and the RMSE was 0.0135. The predicted value exceeded the epidemic threshold on November 21, 2017, by this model. KCDC announced the influenza epidemic alert on December 1, 2017.

Figure 3. A log-scaled plot. Linear regression between F(t) and influenza-like illness (ILI) visit proportion of children younger than 7 years during the development period. Black dots show actual values, and red line shows the result of linear regression.

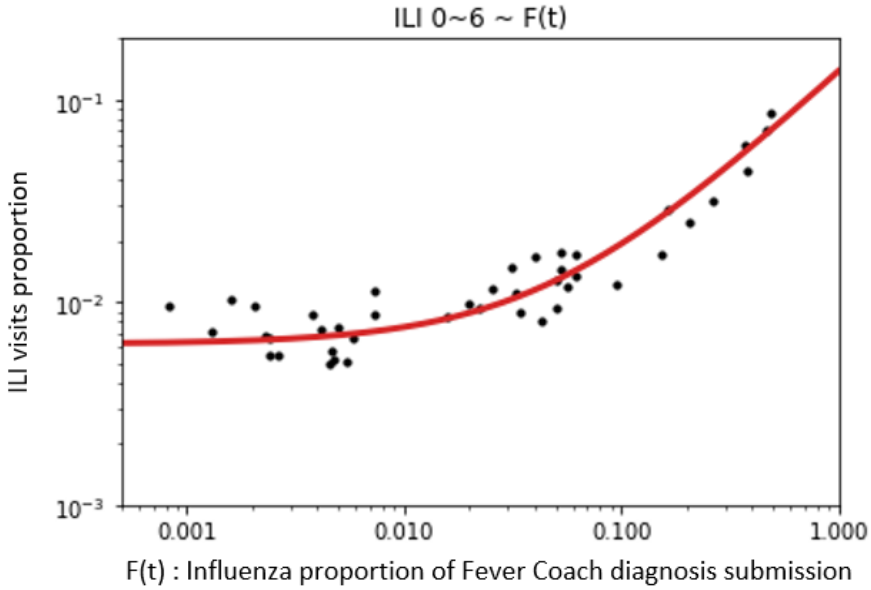
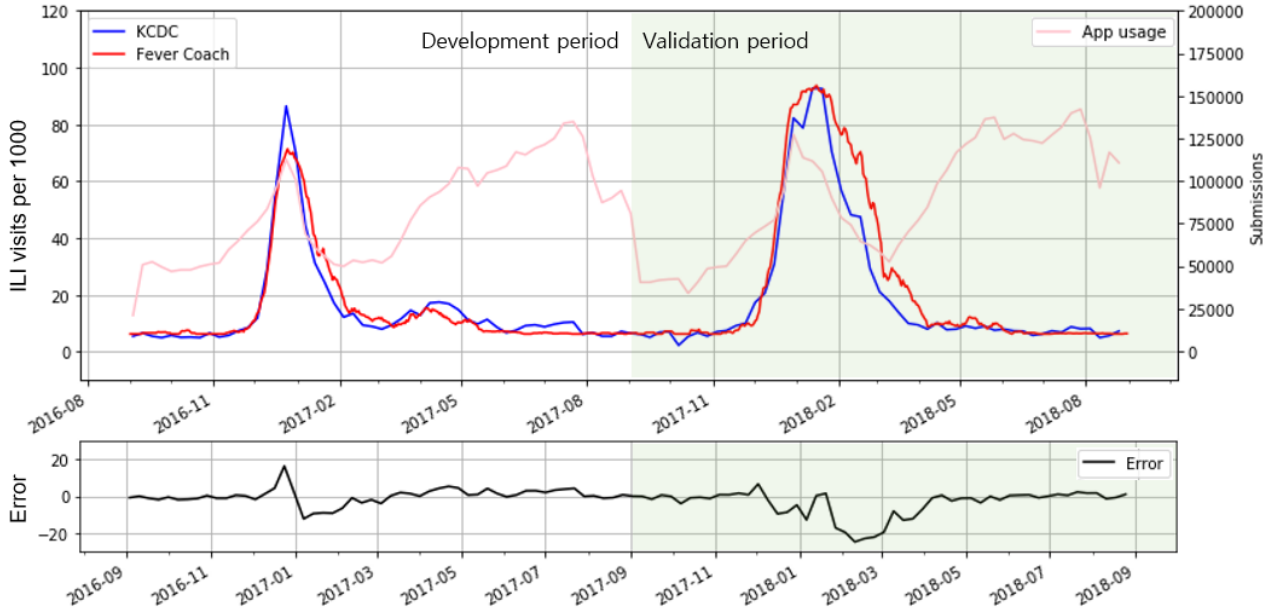


Figure 4. A comparison between Korea Centers for Disease Control and Prevention (KCDC) influenza-like illness (ILI) data and Fever Coach data: blue line shows ILI visits per 1000 outpatient visits reported by KCDC, red line shows ILI prediction from Fever Coach data, and pink line shows the total number of fever-related records in a week. Errors between P (t) and predicted values were displayed at the bottom plot as a black line. The development period was illustrated with a white background color, and green-colored background was used for the validation period.



Discussion

Principal Findings

The Fever Coach app successfully collected data from 7.7% of the target population over a 1-year period, which is remarkably higher than participation rates of previously reported Web- or mobile-based participatory surveillance systems that reported participation rates between 0.02% [24] and 0.13% [25]. The influenza activity monitoring model based on voluntary data submitted through the app showed a close association with the large-scale government-reported influenza activity. The participatory data are updated daily, rather than weekly, and the prediction model based on these data led to more sensitive detection of an influenza epidemic compared with the epidemic alert from the KCDC.

The result of this study sheds light on the potential of mobile apps in collecting data for public health surveillance and highlights the importance of user benefits for voluntary data submission. Previous participatory systems for influenza surveillance relied on the social responsibility of individuals for recruiting and retention of participants, which resulted in low participation rate and biased representation of target populations.

Comparison With Previous Work—Recruitment and Retention

There are several other Web- or mobile phone-based survey platforms for participatory surveillance of influenza and other infectious diseases [17,26,24,27]. Some only collect information regarding symptoms [24] and health-related behavior, whereas others also ask outcomes of medical services [17]. These platforms rely on volunteers by advertising the benefits of self-reported data in disease surveillance and do not incentivize individual participants. Therefore, recruitment and retention of users have been identified as the main challenges for them [24]. They used social media, national television, institutional networks, professional societies, dissemination events, and word of mouth to recruit participants. Using this strategy, Flu Near You (FNY) in the United States recruited 61,000 participants during its second influenza season (2012-2013), which was about 24 per 100,000 population of age 13 years and older. Influenzanet, which was established in 5 European countries in 2009, reported 36,192 participants in 10 countries in Europe during the 2015-2016 influenza season, with a participation rate of 13 individuals per 100,000 population.

In South Korea, online communities are very active and Koreans seek information and communicate with their peers about studying, career, hobbies, and parenting through such communities as well as blogs [28]. Fever Coach used these communities to promote the app by providing free coffee coupons for randomly selected users who downloaded the app and shared their experience and by providing free online health consultation. After reaching 100,000 downloads in May 2016, there have not been active marketing activities, but parents who used the app have voluntarily spread the word on the internet and in their local communities. In February 2017, Fever Coach had been downloaded 300,000 times and over 200,000 users submitted their data at least once. Using this strategy, the participation rate was about 7700 per 100,000 population aged 1 year to 6 years, one of the high-risk groups for influenza infection. The Fever Coach app was primarily designed to enable parents to take care of their kids when they have any kind of fever, but it was able to collect data that can be used specifically for influenza surveillance. We believe that by providing immediate benefit and targeting the vulnerable population, recruiting and retention of participatory surveillance can be much more effective than by relying on volunteers. This strategy

can be applied to target other high-risk groups such as pregnant women, elderly people, or immunocompromised individuals.

Comparison With Previous Work—User Characteristics

As users of Fever Coach submit data on behalf of their children, the only information we collected about the users themselves was email address and device location. However, as the service is designed for caregivers of children younger than 5 years, we can assume that the users' age range is approximately between late 20s and early 40s, which is younger than the user age ranges reported by other systems. As the app is only accessible via a smartphone, every user is assumed to be a smartphone user. The rate of smartphone ownership in South Korea is reported to be as high as 89.5% [29,30], and the smartphone usage rate is above 99% in the age group between 20 and 50 years [29]. Given the very high rate of smartphone penetration, smartphone users represent the general population of this age group in South Korea. Although there are no data about the gender of the users, it is expected that more females than males use the app based on the higher dependency on mothers for childcare in South Korea. This aligns with other studies that reported female participants being the majority. In a study that analyzed data from FNY, participants reporting for household members showed higher participation rate. This finding can partially explain the successful user recruitment and retention of Fever Coach, which is mostly used by parents for their children. As the service itself is targeted for children, the data collected from the app can only represent the pediatric population. However, our model from these data was able to predict influenza activity in the total population as well as in the pediatric population. This was because influenza activity in children is very closely related to that in adolescents and adults, but it may not be applicable for monitoring other diseases. Another noteworthy finding was the close prediction of influenza activity using the prediction model developed from the previous year despite several changes in the user population. Compared with the development period, the number of users decreased and the median number of data submission per child increased. This finding suggests that retention of active users can make up for the loss of inactive users for participatory surveillance.

Limitations

There are many limitations of this study that need to be addressed. First, most of the data submitted to Fever Coach came from children with febrile illness, and users generally do not use the app when their children are well. Therefore, increased app usage can either indicate an outbreak of an infectious disease or a surge of interest in the app itself following national media coverage. To overcome this limitation, we only used *diagnosis reports* to estimate ILI visits, contrary to other participatory surveillance systems that ask symptoms of users and use the data for syndromic surveillance. Although this strategy led to a good approximation of influenza activity, it introduces limited generalization. In most regions of South Korea, with physician density of 2.3 per 1000 population [31], primary care is considered affordable and readily accessible. However, in areas where health care is not as accessible, use of physician-made diagnosis may result in a bias toward a small group of people who have better access to health care than the general population. To overcome this second limitation, an additional study to develop more complex prediction models that include symptom and demographic data is in progress.

Conclusions

The Fever Coach app successfully collected data from 7.73% (207,699/2,686,580) of the target population by providing care instruction for febrile children. These data were used to develop a model that accurately estimated influenza activity measured by the central government agency using reports from sentinel facilities in the national surveillance network.

Acknowledgments

Namsoo Oh, CEO of Mobile Doctor, and Jaewon Shin, CEO of Mobile Doctor, provided the data used for this study. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, Information and Communication Technology, and Future Planning (NRF-2017R1E1A1A03070934). This study used data from Mobile Doctor as well as Department of infectious disease surveillance, KCDC.

Authors' Contributions

HWH led the study and the discussion and review for the entire study. MK developed and designed the study, analyzed the data, and wrote the manuscript. SY reviewed the literature, designed the study, and wrote the manuscript. YJ and SOS managed the project and reviewed the manuscript. SC designed the study and analyzed the data. The final manuscript was approved by all the authors. HWH is the guarantor.

Conflicts of Interest

MK is a CMIO of Mobile Doctor which makes the service Fever Coach (in Korean) and he has equity in Mobile Doctor less than 5%. SC is a CTO of Mobile Doctor and he has equity in Mobile Doctor less than 5%. No other author has reported any other conflict of interest relevant to this study.

Multimedia Appendix 1
A video of visualization. Visualization shows diagnosis report submission on the map geographically as time passes.

[MP4 File (MP4 Video), 25544 KB - mhealth_v7i10e14276_app1.mp4]

References

1. Thompson WW, Weintraub E, Dhankhar P, Cheng P, Brammer L, Meltzer MI, et al. Estimates of US influenza-associated deaths made using four different methods. Influenza Other Respir Viruses 2009 Jan;3(1):37-49 [FREE Full text] [doi: 10.1111/j.1750-2659.2009.00073.x] [Medline: 19453440]
2. World Health Organization. 2018. Influenza (Seasonal) URL: https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal) [accessed 2019-01-30]
3. Nair H, Brooks WA, Katz M, Roca A, Berkley JA, Madhi SA, et al. Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. Lancet 2011 Dec 3;378(9807):1917-1930. [doi: 10.1016/S0140-6736(11)61051-9] [Medline: 22078723]
4. World Health Organization. 2017. Global Influenza Surveillance and Response System (GISRS) URL: https://www.who.int/influenza/gisrs_laboratory/en/ [accessed 2019-01-30]
5. World Health Organization. 2017. National Influenza Centres URL: https://www.who.int/influenza/gisrs_laboratory/national_influenza_centres/national_influenza_centres.pdf?ua=1 [accessed 2019-01-30]
6. Steffen C, Debellut F, Gessner BD, Kasolo FC, Yahaya AA, Ayebazibwe N, et al. Improving influenza surveillance in sub-Saharan Africa. Bull World Health Organ 2012 Apr 1;90(4):301-305 [FREE Full text] [doi: 10.2471/BLT.11.098244] [Medline: 22511827]
7. Ortiz JR, Sotomayor V, Uez OC, Oliva O, Bettels D, McCarron M, et al. Strategy to enhance influenza surveillance worldwide. Emerg Infect Dis 2009 Aug;15(8):1271-1278 [FREE Full text] [doi: 10.3201/eid1508.081422] [Medline: 19751590]
8. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
9. Choi J, Cho Y, Shim E, Woo H. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. BMC Public Health 2016 Dec 8;16(1):1238 [FREE Full text] [doi: 10.1186/s12889-016-3893-0] [Medline: 27931204]
10. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
11. Santillana M, Nsoesie EO, Mekaru SR, Scales D, Brownstein JS. Using clinicians' search query data to monitor influenza epidemics. Clin Infect Dis 2014 Nov 15;59(10):1446-1450 [FREE Full text] [doi: 10.1093/cid/ciu647] [Medline: 25115873]
12. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: a comparative analysis. JMIR Public Health Surveill 2016 Oct 20;2(2):e161 [FREE Full text] [doi: 10.2196/publichealth.5901] [Medline: 27765731]
13. Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating influenza outbreaks using both search engine query data and social media data in south Korea. J Med Internet Res 2016 Jul 4;18(7):e177 [FREE Full text] [doi: 10.2196/jmir.4955] [Medline: 27377323]
14. Kandula S, Hsu D, Shaman J. Subregional nowcasts of seasonal influenza using search trends. J Med Internet Res 2017 Nov 6;19(11):e370 [FREE Full text] [doi: 10.2196/jmir.7486] [Medline: 29109069]
15. Yang S, Santillana M, Brownstein JS, Gray J, Richardson S, Kou SC. Using electronic health records and internet search information for accurate influenza forecasting. BMC Infect Dis 2017 May 8;17(1):332 [FREE Full text] [doi: 10.1186/s12879-017-2424-7] [Medline: 28482810]
16. Smolinski MS, Crawley AW, Olsen JM, Jayaraman T, Libel M. Participatory disease surveillance: engaging communities directly in reporting, monitoring, and responding to health threats. JMIR Public Health Surveill 2017 Oct 11;3(4):e62 [FREE Full text] [doi: 10.2196/publichealth.7540] [Medline: 29021131]
17. Koppeschaar CE, Colizza V, Guerrisi C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: citizens among 10 countries collaborating to monitor influenza in Europe. JMIR Public Health Surveill 2017 Sep 19;3(3):e66 [FREE Full text] [doi: 10.2196/publichealth.7429] [Medline: 28928112]
18. Bajardi P, Paolotti D, Vespignani A, Eames K, Funk S, Edmunds WJ, et al. Association between recruitment methods and attrition in internet-based studies. PLoS One 2014;9(12):e114925 [FREE Full text] [doi: 10.1371/journal.pone.0114925] [Medline: 25490045]
19. Kim JH, Yun S, Hwang SS, Shim JO, Chae HW, Lee YJ, Committee for the Development of Growth Standards for Korean Children and Adolescents, Committee for School Health and Public Health Statistics, the Korean Pediatric Society, Division of Health and Nutrition Survey, Korea Centers for Disease Control and Prevention. The 2017 Korean National Growth Charts for children and adolescents: development, improvement, and prospects. Korean J Pediatr 2018 May;61(5):135-149 [FREE Full text] [doi: 10.3345/kjp.2018.61.5.135] [Medline: 29853938]
20. Centers for Disease Control and Prevention Korea. 2019. Stastical Information of Sentinel Surveillance of Influenza, 2019 URL: http://www.cdc.go.kr/npt/biz/npp/iss/influenzaStatisticsMain.do [accessed 2019-01-30]
21. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2015;12(2011):2825-2830 [FREE Full text]

22. National Statistics Portal - KOSIS. 2019. Summary of Census Population (By Administrative District/Sex/Age), 2017 URL: http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1IN1503&conn_path=I2 [accessed 2019-01-30]
23. Fever Coach. 2019. Fever Coach Flu Visualization URL: http://ec2-3-83-42-212.compute-1.amazonaws.com/ [accessed 2019-01-30]
24. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/AJPH.2015.302696] [Medline: 26270299]
25. Dalton C, Carlson S, Butler M, Cassano D, Clarke S, Fejsa J, et al. Insights from flutracking: thirteen tips to growing a web-based participatory surveillance system. JMIR Public Health Surveill 2017 Aug 17;3(3):e48 [FREE Full text] [doi: 10.2196/publichealth.7333] [Medline: 28818817]
26. Fujibayashi K, Takahashi H, Tanei M, Uehara Y, Yokokawa H, Naito T. A new influenza-tracking smartphone app (flu-report) based on a self-administered questionnaire: cross-sectional study. JMIR Mhealth Uhealth 2018 Jun 6;6(6):e136 [FREE Full text] [doi: 10.2196/mhealth.9834] [Medline: 29875082]
27. Chan TC, Hu TH, Hwang JS. Estimating the risk of influenza-like illness transmission through social contacts: web-based participatory cohort study. JMIR Public Health Surveill 2018 Apr 9;4(2):e40 [FREE Full text] [doi: 10.2196/publichealth.8874] [Medline: 29631987]
28. Nguyen T. Medium. 2017. Online Communities in South Korea: A Case Study URL: https://medium.com/@thinguyens/online-communities-in-south-korea-a-case-study-5815dd589630 [accessed 2019-01-30]
29. Korea Internet Information Center (KRNIC). 2018. Survey on the Internet Usage, 2017 URL: https://isis.kisa.or.kr/board/index.jsp?pageId=060100&bbsId=7&itemId=821&pageIndex=1 [accessed 2019-01-30]
30. Poushter J. Pew Research Center. 2016. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2019-01-15]
31. World Health Organization. 2018. Global Health Observatory Data Repository URL: http://apps.who.int/gho/data/node.main.A1444 [accessed 2019-01-15]

Abbreviations

FNY: Flu Near You
GFT: Google Flu Trends
ILI: influenza-like illness
IRB: institutional review board
KCDC: Korea Centers for Disease Control and Prevention
NRF: National Research Foundation of Korea
RMSE: root mean square error

Edited by G Eysenbach; submitted 06.04.19; peer-reviewed by D Paolotti, L Sheets; comments to author 28.05.19; revised version received 14.07.19; accepted 09.08.19; published 17.10.19.

Please cite as:
Kim M, Yune S, Chang S, Jung Y, Sa SO, Han HW
The Fever Coach Mobile App for Participatory Influenza Surveillance in Children: Usability Study
JMIR Mhealth Uhealth 2019;7(10):e14276
URL: https://mhealth.jmir.org/2019/10/e14276
doi:10.2196/14276
PMID:31625946

Original Paper

Usability and Feasibility of a Smartphone App to Assess Human Behavioral Factors Associated with Tick Exposure (The Tick App): Quantitative and Qualitative Study

Maria P Fernandez[1,2*], PhD; Gebbiena M. Bron[3*], DVM, PhD; Pallavi A Kache[2], MPH; Scott R Larson[3], PhD; Adam Maus[4], MSc; David Gustafson Jr[4], MSc; Jean I Tsao[5], PhD; Lyric C Bartholomay[6], PhD; Susan M Paskewitz[3], PhD; Maria A Diuk-Wasser[2], PhD

[1]Earth Institute, Columbia University, New York City, NY, United States

[2]Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, United States

[3]Department of Entomology, University of Wisconsin-Madison, Madison, WI, United States

[4]Center for Health Enhancement System Studies, University of Wisconsin-Madison, Madison, WI, United States

[5]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, United States

[6]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, United States

[*]these authors contributed equally

Corresponding Author:
Maria A Diuk-Wasser, PhD
Department of Ecology, Evolution and Environmental Biology
Columbia University
Schemerhorn Ext Building, 11th Floor, Room 1013
New York, NY
United States
Phone: 1 212 854 3355
Email: mad2256@columbia.edu

Abstract

Background: Mobile health (mHealth) technology takes advantage of smartphone features to turn them into research tools, with the potential to reach a larger section of the population in a cost-effective manner, compared with traditional epidemiological methods. Although mHealth apps have been widely implemented in chronic diseases and psychology, their potential use in the research of vector-borne diseases has not yet been fully exploited.

Objective: This study aimed to assess the usability and feasibility of The Tick App, the first tick research–focused app in the United States.

Methods: The Tick App was designed as a survey tool to collect data on human behaviors and movements associated with tick exposure while engaging users in tick identification and reporting. It consists of an enrollment survey to identify general risk factors, daily surveys to collect data on human activities and tick encounters (Tick Diaries), a survey to enter the details of tick encounters coupled with tick identification services provided by the research team (Report a Tick), and educational material. Using quantitative and qualitative methods, we evaluated the enrollment strategy (passive vs active), the user profile, location, longitudinal use of its features, and users' feedback.

Results: Between May and September 2018, 1468 adult users enrolled in the app. The Tick App users were equally represented across genders and evenly distributed across age groups. Most users owned a pet (65.94%, 962/1459; P<.001), did frequent outdoor activities (recreational or peridomestic; 75.24%, 1094/1454; P<.001 and 64.58%, 941/1457; P<.001, respectively), and lived in the Midwest (56.55%, 824/1457) and Northeast (33.0%, 481/1457) regions in the United States, more specifically in Wisconsin, southern New York, and New Jersey. Users lived more frequently in high-incidence counties for Lyme disease (incidence rate ratio [IRR] 3.5, 95% CI 1.8-7.2; P<.001) and in counties with cases recently increasing (IRR 1.8, 95% CI 1.1-3.2; P=.03). Recurring users (49.25%, 723/1468) had a similar demographic profile to all users but participated in outdoor activities more frequently (80.5%, 575/714; P<.01). The number of Tick Diaries submitted per user (median 2, interquartile range [IQR] 1-11) was higher for older age groups (aged >55 years; IRR 3.4, 95% CI 1.5-7.6; P<.001) and lower in the Northeast (IRR[NE]

0.4, 95% CI 0.3-0.7; *P*<.001), whereas the number of tick reports (median 1, IQR 1-2) increased with the frequency of outdoor activities (IRR 1.5, 95% CI 1.3-1.8; *P*<.001).

Conclusions: This assessment allowed us to identify what fraction of the population used The Tick App and how it was used during a pilot phase. This information will be used to improve future iterations of The Tick App and tailor potential tick prevention interventions to the users' characteristics.

(JMIR Mhealth Uhealth 2019;7(10):e14769) doi:10.2196/14769

KEYWORDS

Lyme disease; ticks; ecological momentary assessment; citizen science

Introduction

The Use of Mobile Health in Vector-Borne Diseases

In the United States, the number of adults who own a smartphone has been steadily increasing over time for every age group within the adult population. It is estimated that 81% of the adult population owns a smartphone compared with 15% of the population owning a normal mobile phone [1]. The ubiquity of smartphones provides a unique opportunity to gather and share information about health and disease, given the capabilities such as location sensing and software apps. As a result, mobile health (mHealth) technology is becoming an important part of health care service provision and is changing the way in which people use health information and communicate with health organizations and health professionals [2,3]. mHealth allows the general public to access a health service where and when they need it. In the context of public health, mHealth is particularly suited for patient education, disease self-management, and remote monitoring of patients [4]. Moreover, the use of mHealth can leverage smartphone features to turn them into research tools with the potential to reach a larger section of the population in a more cost-effective manner than traditional epidemiological methods and transform survey instruments into high-frequency (fine temporal resolution), spatially resolved data collection tools [5]. The widespread use of smartphone apps can be thought of as a 2-way communication channel between affected users and researchers. Although apps have been widely implemented in chronic diseases and psychology [4], their potential use in the research of vector-borne diseases has not yet been fully exploited. Most mHealth apps applied to vector-borne disease research have targeted mosquito-borne diseases [6-9] and only a few have targeted tick vectors. These are limited to the country level in Europe [10,11].

Lyme Disease Risk

Lyme disease is the most commonly reported vector-borne disease in the United States, with 300,000 cases estimated per year [12,13], the majority of which are reported from the northeastern and north-central states [12]. In these areas, Lyme disease risk is determined by human exposure to infected *Ixodes scapularis* ticks, which can occur either peridomestically or within natural areas [13-18]. The seasonality of human cases mirrors that of *I. scapularis* nymphal activity; nymphs are abundantly active from May to early August, peaking in early-mid summer [12]. The association between human cases and nymphal activity can be in part attributed to the small size of nymphal ticks compared with adults, resulting in prolonged or undetected attachment. Human exposure to ticks depends on the density of infected ticks, but this association is strongly modified by local conditions, including human behavior [19,20]. Human behaviors that have been shown to influence tick exposure include the frequency and type of outdoor activities and adaptive responses following interactions with tick habitat [13,21,22]. In turn, prior exposure to ticks may trigger multiple behavioral responses to reduce exposure, such as avoidance of tick habitat and the use of personal protection measures [16].

The links between human activity, mobility patterns, and tick exposure, however, have not been well documented in Lyme disease–endemic areas of the United States, in part, because of methodological limitations. Traditional mechanisms of data collection of human behavior (eg, retrospective questionnaires) are subject to uncertainties in the degree to which findings can be generalized beyond the investigation depending on how and where they are administered [23]; surveys administered through an app can help expand the population surveyed at a lower cost. In addition, survey results are often affected by recall bias and do not capture specific intrasubject variability, particularly when participants are required to generalize an experience or behavior [23]. To address these issues, we developed a smartphone app, The Tick App [24], to (1) serve as a research tool to better understand human behaviors affecting tick exposure and (2) engage the general public in active tick prevention and reporting in different regions of the United States. As a research tool, it includes epidemiological surveys using *ecological momentary assessments* (EMA) to gather quantitative behavioral data within the context of a participant's daily routine [25]. There is an increasing body of literature describing the use of smartphone app technology as a platform for EMA implementation, particularly within psychology and substance abuse research [23,26,27]. To our knowledge, however, this method has not been leveraged in assessing behavioral risk factors for Lyme disease.

Objective

In this study, we report on the development and implementation of The Tick App in the United States. Our overall aim was to assess its usability and feasibility and identify what fraction of the population engaged in The Tick App to better interpret the external validity of the app-derived data. Information from this study will be used to improve future iterations of The Tick App and tailor potential tick prevention interventions to the user population.

Methods

Design of The Tick App (Phases 1 and 2)

The Tick App was developed from a preprototype named *GeoQuestion*, which was designed and evaluated during the spring and summer of 2017 (phase 1: preprototype [28]; Table 1). On the basis of this experience, we developed The Tick App (phase 2: prototype [28]; Multimedia Appendix 1). To align the design of the app with potential users' characteristics, we framed the activities based on a roadmap for creating mHealth apps [29] (Table 1). Coding of The Tick App was led by the Center for Health Enhancement System Studies at the University of Wisconsin-Madison, and a prototype was completed by April 2018. After testing the prototype extensively, the app was available for download in May 2018 (Multimedia Appendix 1). The app is available for Android and iOS operating systems and can be downloaded from smartphone app stores at no cost (at the time of this publication). This study complied with the Consolidated Standards of Reporting Trials (Multimedia Appendix 2).

Table 1. Phases, objectives, activities, and timeline in the design process of The Tick App based on the roadmap proposed in a study by van Velsen et al for creating mobile health apps and the guidelines for monitoring and evaluating mobile health interventions developed by the World Health Organization.

Phases of development [29]	Objective	Activities	Timeline	World Health Organization phase [28]
Contextual inquiry	Identification of end users and the context in which the app will be used	Identify research objectives; focus groups with the users of GeoQuestion preprototype app	September to November 2017	Phase 1: preprototype
Value specification	Identification of end users' values and requirements from phase 1	Design the content of The Tick App	December 2017 to January 2018	Phase 2: prototype design
Design	Creation and testing of prototype	Code The Tick App; pilot testing of prototype; focus groups in target study areas	February to April 2018	Phase 2: prototype design
Operationalization	Launch and recruitment of the app	Passive and active recruitment of users; collect data generated by The Tick App users	May to August 2018	Phase 3a: pilot the prototype (usability)
Summative evaluation	Gather feedback from The Tick App users	Focus groups with The Tick App users	September 2018	Phase 3b: pilot the prototype

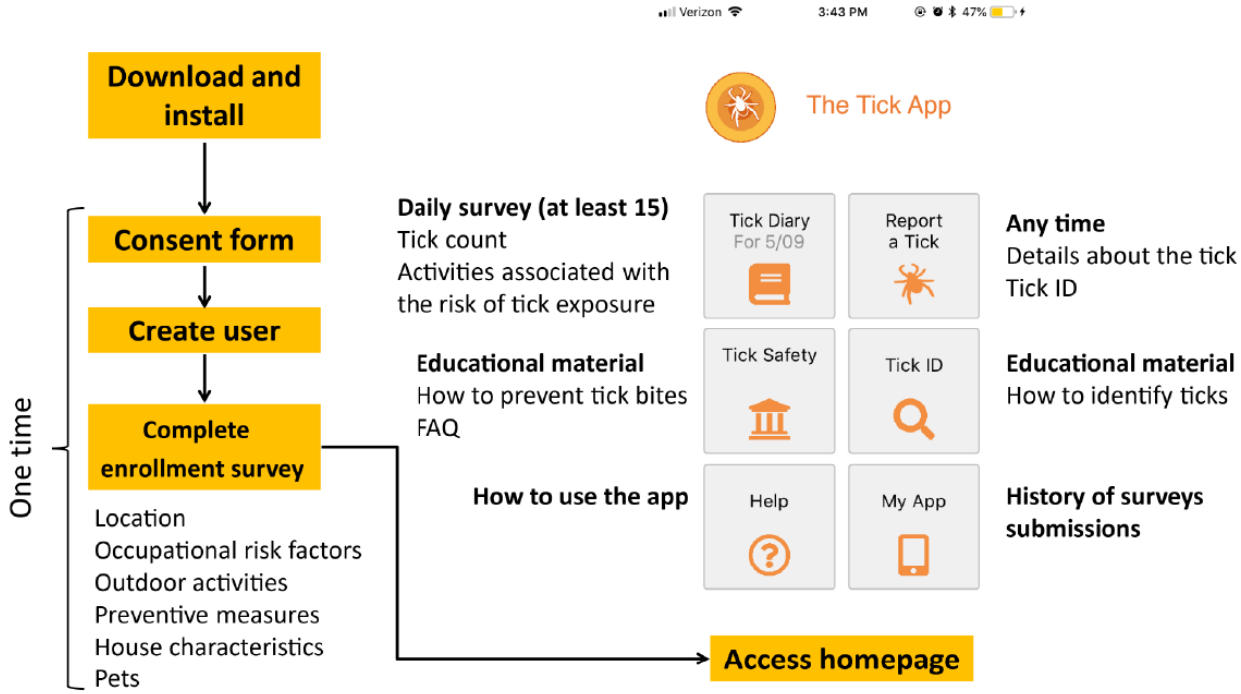
Functionalities and Workflow of The Tick App

After users downloaded and installed The Tick App, a brief explanation of the study and electronic informed consent were displayed (Figure 1). The informed consent had to be accepted before proceeding to create a user account. The Tick App functionalities included an enrollment survey, which was modified from the preprototype. This survey was designed to take less than 10 min to fill out and aimed to collect the users' demographic data, house characteristics, past experiences with ticks and tick-borne diseases, typical use of personal prevention methods and household interventions to reduce tick encounters, and general frequency of outdoor and peridomestic activities during the spring and summer. This survey was completed only once by the user before accessing the homepage.

On the homepage, several functionalities were available at any time of the day. In contrast, the *Tick Diary* was only available once a day (from 5 pm to 10 am CST; Figure 1). The *Tick Diary* was a daily retrospective survey that collected information on the user's daily outdoor activities, tick encounters on themselves, their pets or members of their household, and any personal protection measures used to prevent tick bites. This survey collected fine-scale temporal data about human behavior in association with tick encounters, and it was designed to take less than 1 min to complete. We asked all users to submit at least 15 daily surveys to gain representation of the typical activity patterns for each individual, assess the variability between users, and calculate an accurate measure of the risk of tick encounter. A pop-up notification showed up every day at 5 pm CST until 15 *Tick Diaries* were completed. This study follows the recommended guidelines for reporting results of internet e-surveys (Multimedia Appendix 3).

Users were also able to report any ticks they found through the *Report a Tick* survey in the app and could send a picture of the tick via a Web-based form external to the app. These tick reports collected information on the tick encounter: who had the tick (ie, self, pet, or member of the household), if it was attached, and where the exposure might have occurred. Users were also asked to identify the tick from photographs that were provided in the app, including photos of female and male adult *Dermacentor variabilis*, *Amblyomma americanum*, and *I scapularis* as well as an *I scapularis* nymph. If users sent a picture of the tick via the online survey, trained members of our research group with experience on taxonomic identification of ticks identified the tick to the species and life stage visually from the picture and provided this information to the user via email. The confidence in the identification depended on the quality of the picture submitted and varied between stages (easier in adults compared with nymphs). On the basis of our experience, we were able to identify 94.8% (502/533) of pictures submitted to the species level with a high degree of confidence in the identification.

Figure 1. Workflow and homepage of The Tick App. The time frequency for the surveys (Tick Diary and Report a Tick) and the type of content of the remaining functionalities displayed on the homepage were indicated. FAQ: frequently asked question; ID: identification.



Through the app, we also provided educational resources about the biology and ecology of ticks (*FAQ* and within *Tick Safety*), how to identify them (*Tick ID*), and measures for protection against them (*Tick Safety*). All functionalities were accessible offline, including the surveys (*Tick Diary* and *Report a Tick*), which were stored locally on the smartphone, but needed an online connection to be submitted. Information about the number of surveys submitted, log out button, and the informed consent were available in *My App*. The *Help* functionality explained the use of the app to the users.

The Tick App was also able to use location services through the smartphone to capture the global positioning system (GPS) coordinates if the user allowed for this functionality. If location services were enabled, the app would record GPS coordinates every 15 min depending on the cell phone service. A more detailed explanation of the backend components of The Tick App and the surveys' design can be found in the technical appendix (Multimedia Appendix 4).

Usability of The Tick App (Phase 3a)

We evaluated the usability of The Tick App from April 20 to September 3, 2018 (phase 3a [28]; Table 1). We assessed our dissemination and recruitment strategies, the profile of The Tick App users with respect to their demographics, risk factors associated with tick exposure risk, and geographical distribution. We also evaluated the longitudinal use of The Tick App and how users engaged with different app functionalities.

User Enrollment

The Tick App was available to any adult aged older than 18 years who lived in the United States, owned a smartphone, and was familiar with software apps. Active and passive recruitment of users began with a special focus on the Midwest and Northeast, particularly in Wisconsin and southern New York state, respectively. At these locations, passive recruitment occurred by media coverage. Active recruitment was undertaken during house visits coupled with ongoing field research involving tick sampling in yards at selected study sites (Eau Claire, WI, and Staten Island, NY). During these visits, the researchers explained the objective of the app and invited residents to participate as users (Multimedia Appendix 5). Recruitment was conducted during June and July 2018. All activities complied with the ethical principles included in the Declaration of Helsinki as reviewed by Institutional Review Boards (IRBs) of Columbia University (IRB protocol: AAA3750-M00Y01) and the University of Wisconsin-Madison (IRB protocol: 2018-84). Participants provided their informed consent through the app. We assessed the longitudinal enrollment pattern during the study period and the effects of the active recruitment and media coverage (passive recruitment).

User Profile

To characterize the user profile, we considered demographic variables (age and gender), risk factors associated with tick encounters (type of property, owning a pet, and frequency of outdoor activities), and previous experience with ticks and tick-borne diseases. *Frequent peridomestic activity* described users who reported mowing the lawn or gardening once a week. *Frequent outdoor recreation* described users who went hiking, running, or biking on nature trails or visited the beach (at a lake or ocean) at least once a week or went hunting, fishing, bird watching, or camping at least once a month during the spring and summer months.

We used multiple correspondence analysis (MCA) as an exploratory analysis of the association among different user characteristics [30]. The MCA is the equivalent of principal component analysis but for categorical variables; it reduces the dimensionality of the covariance matrix into linear combinations

of the original variables (dimensions) and decomposes the variance (inertia) of the sample [30]. The different dimensions can be assessed graphically using biplots, which allow a better understanding of how the variables are interrelated and their relative contribution to the score [31]. We used a first MCA including all users' characteristics (age, gender, owning a pet, type of house, frequent outdoor recreation, outdoor work or volunteer job, and frequent peridomestic activity) to assess the association between the variables. We conducted a second MCA, including only owning a pet, frequent outdoor recreation, outdoor work or volunteer job, and frequent peridomestic activity, to construct a summary index of the frequency of outdoor activities (ie, the outdoor activity index). As the first dimension captures most of the inertia, the coordinate (value estimated for each individual based on their characteristics and which represents the contribution to the inertia in the study population) can be used as a quantitative index [32].

Geographic Distribution of Users

When assessing the geographic distribution of users in the United States, we considered the major census regions used by the US Census Bureau [33]. We evaluated if the user's profile varied by region and the effects of the Lyme disease reporting status by county on the number of users. We used Lyme disease case data publicly available from the Centers for Disease Control and Prevention (CDC) [34] to estimate Lyme disease incidence at a county level. We used the number of cases reported (*confirmed* and *probable* cases) per county for a 5-year period (2013-2017) and the population size obtained from the National Census in 2010 to estimate a 5-year period Lyme disease annual incidence per county and percentage change in cases within that period. The 5-year period Lyme disease annual incidence was estimated as follows: (cumulative number of cases between 2013 and 2017)/(population size × 5 years) ×100,000. The mid-year total population was estimated as the population size in 2010 multiplied by the duration of the period [35]. We designated counties to be as *high incidence* if this measure was greater than 10 cases per 100,000, otherwise they were *low-incidence* counties or had *no cases of Lyme disease*. In addition, we used the CDC case data to classify the counties according to their percent change in Lyme disease incidence between 2013 and 2017. By combining this classification with Lyme disease incidence, we further classified the counties into 5 categories: (1) high incidence—no change; (2) high incidence—greater than 1-fold increase; (3) low incidence—no change; (4) low incidence—greater than 1-fold increase; and (5) no Lyme disease cases reported. A greater than 1-fold increase referred to counties where Lyme disease cases at least doubled during 2013 to 2017.

Longitudinal App Use

The data on the longitudinal use of the app were derived from the *Tick Diary*, *Report a Tick*, and screen views. Independent screen views (as opposed to continuous screen views) were considered as screen views at every 5-min intervals, assuming that screen views happening within 4 min or less represented a single interaction with The Tick App. We tested whether the user's profile differed between recurring (ie, interacted with the app beyond the enrollment survey) and nonrecurring users and identified the most frequently used features of the app during the study period. Among recurring users, we evaluated the association between demographic variables, geographic location and outdoor activity pattern, and the number of *Tick Diaries* and tick reports submitted.

Statistical Analysis

To summarize the distribution of numeric variables (eg, number of users), we reported the interquartile range (IQR) as a measure of variability because numeric variables deviated from a normal distribution [36]. For descriptive bivariate analyses of categorical variables, we used chi-square tests and nonparametric Kruskal-Wallis tests when comparing categorical and continuous variables (eg, number of users vs region). We conducted multivariate analyses using generalized linear models [37] and generalized linear mixed models when accounting for random effects (eg, regional effects in the number of users per county) [38]. In the case of binary response variables, we used logistic regression models with logit as the link function and the relative risk expressed as odds ratios (ORs). When the response variable was numeric (count data), we used negative binomial models with log as the link function and the relative risk expressed as incidence rate ratios (IRRs). Negative binomial regression was preferred over Poisson regression, given the overdispersed distributions [39]; both models were compared by log-likelihood ratio tests. All analyses were implemented in Stata version 14.2 [40] and R version 3.2.3 (lme4 and car packages) [41].

Feasibility Evaluation

In September 2018, we organized focus groups in Staten Island, NY, and Eau Claire, WI, by extending personal invitations to local users of The Tick App, with the goal of gathering feedback on the app after its implementation during the study period. In total, 14 users participated in the focus groups. The guiding questions used in the pre- and postimplementation focus groups can be found in Multimedia Appendix 6.

Results

User Enrollment

Between April 20 and September 3, 2018, 1468 (86.09%, 1468/1705) users completed the enrollment survey after downloading the app (Multimedia Appendix 7); 71.53% (1050/1468) users installed the app in smartphones operating with iOS, whereas 28.47% (418/1468) users installed it in smartphones operating with Android. User enrollment in the study mostly occurred from June to mid-July (1187/1468, 80.86% of the total number of users in the study period) and during the weekend (IRR 1.4, 95% CI 1.2-1.5; $P<.001$). Before the official app launch date (May 28, 2018), 81 users enrolled in the study, increasing to 209 within the first week of the official launch and to 521 by the end of the first week of active recruitment (Multimedia Appendix 7). Recruitment during household visits coupled with peridomestic tick sampling produced a 2-fold increase in enrollment on the same day of the visit and the following day, after accounting for weekend and media coverage effects (IRR 1.2, 95% CI 2.0-2.4; $P<.001$). Media coverage of the app also increased users' enrollment by

2-fold on the same day and the day after the coverage (IRR 2.0, 95% CI 1.8-2.2; *P*<.001; Multimedia Appendix 7). The median time to complete the enrollment survey was 6.3 min (IQR 4.7-8.6).

User Profile

The demographic profile of The Tick App users showed equal gender proportions and widespread age distribution, although young adults aged between 18 and 25 years were significantly underrepresented (Table 2). The median age was 48 years (IQR 36-60) and followed a bimodal distribution peaking at 37 years and 55 years (Multimedia Appendix 8). Users were significantly biased toward owning a pet (Table 2), and the majority of users with pets owned at least one dog (797/960, 83.02%). The proportion of users living in houses with yards was significantly higher than those living in apartments, and more than half of the users reported doing frequent peridomestic activities such as gardening and mowing the lawn during the spring and summer (Table 2). Regarding other outdoor activities, users doing at least one outdoor recreational activity once a week in urban parks or natural areas were significantly overrepresented (Table 2). In contrast, 644 out of 1454 users (44.29%) worked or volunteered in outdoor jobs (eg, landscapers and camp counselors), which was significantly less than the proportion who did not (Table 2).

The exploratory MCA, which accounted for 76.2% of the total variability observed within the first 2 dimensions, showed that users reporting frequent outdoor recreational activities were also more frequently doing peridomestic activities at least once a week and working or volunteering outdoors (Figure 2, dimension 1), although the latter was slightly more associated with males than females (Figure 2, dimension 2). These outdoor enthusiasts were also more associated with owning a pet and with living in a house with a yard (Figure 2). Younger adults were associated with living in apartments and less often involved in outdoor activities in general, although they were underrepresented in the total number of users (Figure 2 and Table 2). Dimension 2 of the MCA separated users by their demographic characteristics (age and gender), but it only accounted for 4.3% of the variability, indicating that these variables did not contribute significantly to the differences observed among users. The outdoor activity index derived from the second MCA, which included only the variables that were associated in the first dimension of the previous MCA, accounted for 83.6% of the total variability (Multimedia Appendix 9). This index was significantly associated with a tick encounter in the previous winter or fall (logistic regression: OR 2.2, 95% CI 1.9-2.6; *P*<.001), but not with a previous diagnosis of a tick-borne disease (logistic regression: OR 1.16, 95% CI 0.98-1.38; *P*=.07) after adjusting for age and gender. One-third of users reported having a tick encounter the previous fall or winter when adult ticks are active, and half of the users reported finding a tick on their pet, whereas the percentage of users reporting a previous tick-borne diagnosis was comparatively lower (11.82%, 173/1464; Table 2). Nonetheless, the self-reported cases of Lyme disease in the previous year (2.05% of the users, 30/1461) were still considerably higher than the percentage of confirmed cases in 2017 within the total population from the counties where the majority of users live (median 0.06%, IQR 0.02%-0.10%).

Table 2. Users' profile including demographic variables, type of property, frequent outdoor activities (occupational, recreational, and peridomestic), and previous experience with ticks and Lyme disease, as reported in the enrollment survey.

Variables	Users, n (%)	*P*[a] value	Chi-square (*df*)
Gender (n=1463)		**.85**	**0.0 (1)**
Male	720 (40.05)		
Female	726 (49.46)		
Others/prefer not to say	22 (1.50)		
Age (years; n=1457)[b]		**.001[c]**	**126.0 (5)**
18-24	94 (6.43)		
25-34	265 (18.13)		
35-44	321 (22.96)		
45-54	274 (18.74)		
55-64	319 (21.82)		
≥65	189 (12.93)		
Pet owner (n=1454)		**<.001[c]**	**146.8 (1)**
Yes	962 (65.94)		
No	497 (34.06)		
Type of house (n=1455)		**<.001[c]**	**0.0 (4)**
House with yard	1109 (75.96)		
Apartment	238 (16.30)		
Cabin/cottage	65 (4.45)		
Mobile home	22 (1.51)		
Other	26 (1.78)		
Work or volunteer outdoors (n=1454)		**<.001[c]**	**18.9 (1)**
Yes	646 (44.28)		
No	813 (55.72)		
Frequent outdoor recreation (n=1449)		**<.001[c]**	**368.8 (1)**
Yes	1094 (75.24)		
No	360 (24.76)		
Frequent peridomestic activities (n=1452)		**<.001[c]**	**122.7 (1)**
Yes	941 (64.58)		
No	516 (35.42)		
Tick exposure in the previous fall (n=1458)		**<.001[c]**	**201.5 (1)**
Yes	459 (31.37)		
No	1004 (68.63)		
Previous tick-borne disease diagnosis (n=1459)		**<.001[c]**	**849.0 (1)**
Yes	173 (11.82)		
No	1291 (88.18)		
Tick finding in their pet during the previous fall (N=958)		**.004[d]**	**8.1 (1)**
Yes	525 (54.57)		
No	437 (45.43)		

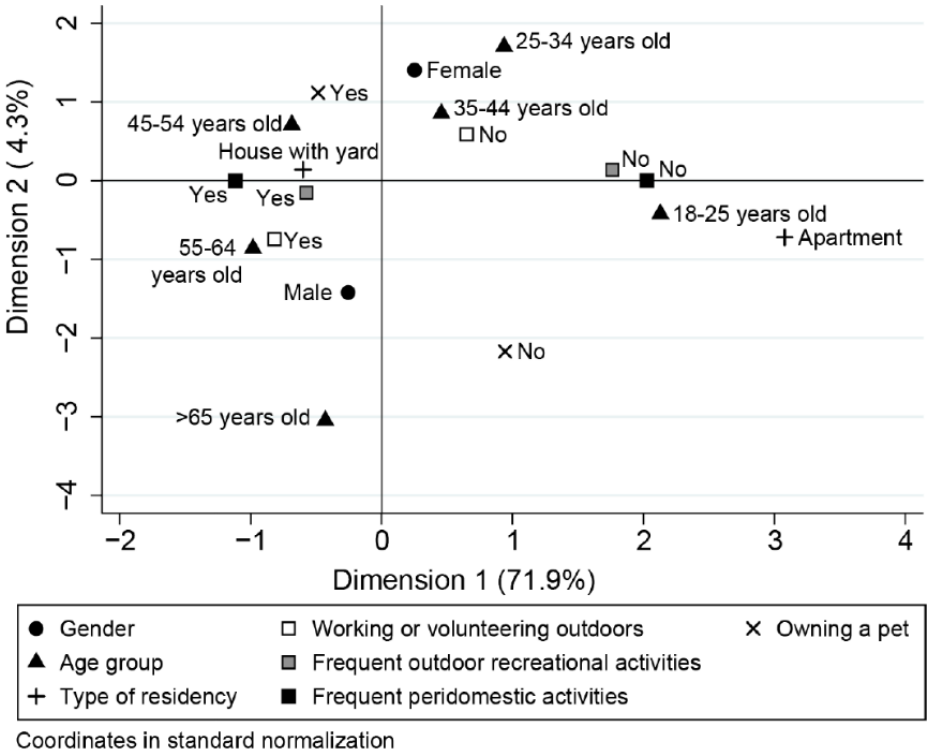
[a]*P* values of the chi-square test of H_0=equal distribution among users are presented.

[b]For age, we tested H_0=equal distribution compared to the total US population (2016 population estimates).

[c]P<.001.

[d].001≤P≤.05.

Figure 2. Biplot of the multiple correspondence analysis (MCA) of the users' profile.



Geographic Distribution of Users

We observed a nationwide distribution of users within the United States, although most of them lived in the Midwest (824/1457, 56.55%), followed by the Northeast (481/1457, 33.01%), and a smaller fraction of users lived in the South and West regions (107/1457, 7.34%, and 45/1457, 3.09%, respectively; Figure 3). Within the Midwest, 82.3% (682/824) of users lived in Wisconsin, whereas users from the Northeast mostly lived in southern New York and New Jersey (305/481, 63.4%, and 65/481, 13.5%, respectively), consistent with the area of influence of our study and recruitment efforts. Most of the users came from high-incidence counties with no recent changes in the number of Lyme disease cases (71.7%, 1045/1457), followed by low-incidence counties (24.9%, 363/1457) and counties with no reported cases (3.4%, 49/1457) (Figure 4). A higher proportion of users in the Northeast downloaded The Tick App on iOS and Android, compared with the Midwest (389/481, 80.9% and 534/824, 64.8%, respectively; χ^2_1=37.8; P<.001). Although the demographic profile was similar across regions, the outdoor activity index differed (Kruskal-Wallis test, $\chi^2_{3=}$93.6; P<.001) and was lower in the Northeast compared with the other regions (Figure 5). Finally, the number of users per county was higher in those with a larger population size and high Lyme disease incidence or with low incidence but a 1-fold increase in the number of reported cases in the previous 5 years (2013-2017), after adjusting for regional random effects (Figure 4 and Table 3).

Figure 3. Number of users per county and per region for the United States by the major census regions in the United States.

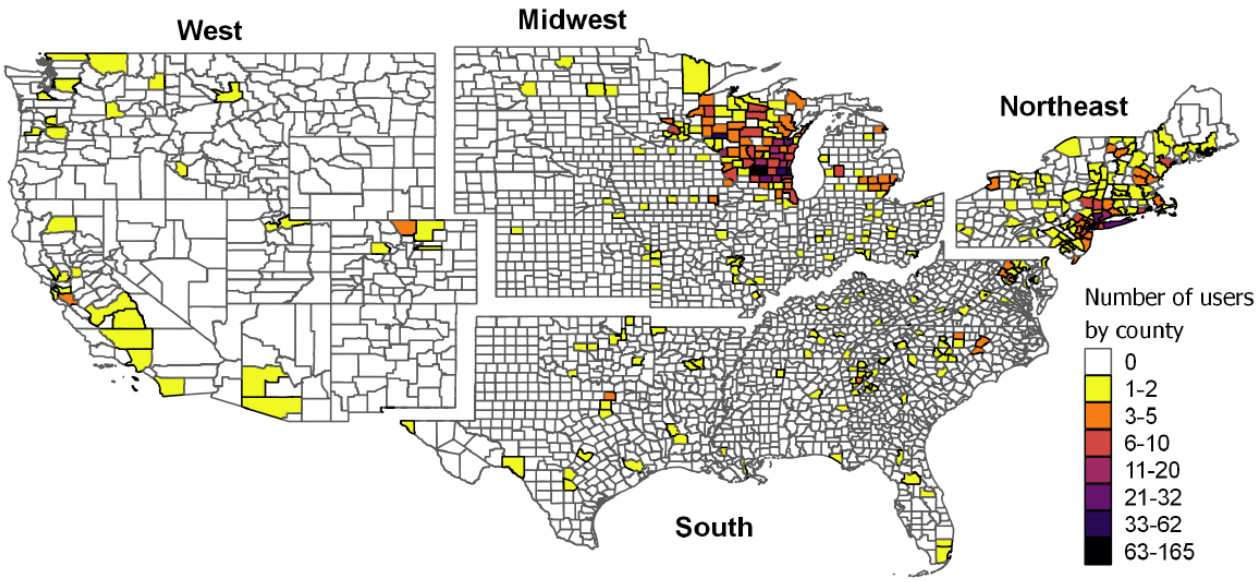


Figure 4. County's Lyme disease status according to Lyme disease incidence and recent increase in 2013-2017.

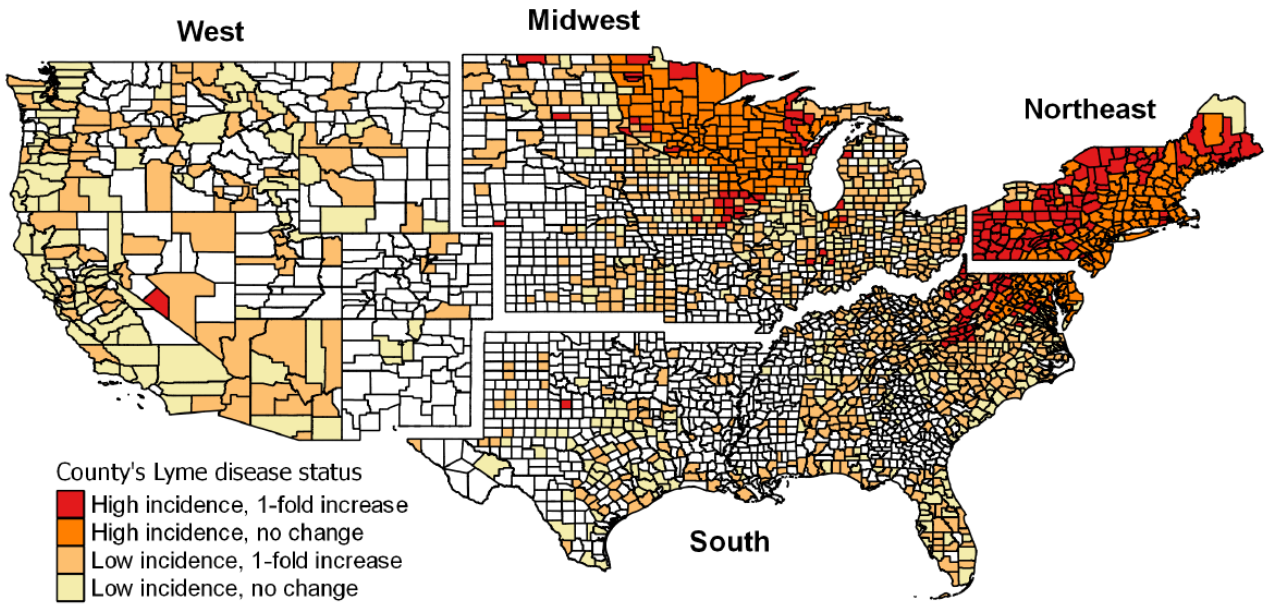


Figure 5. The outdoor activity index derived from the multiple correspondence analysis (MCA) excluding age and gender, by the major census regions in the United States.

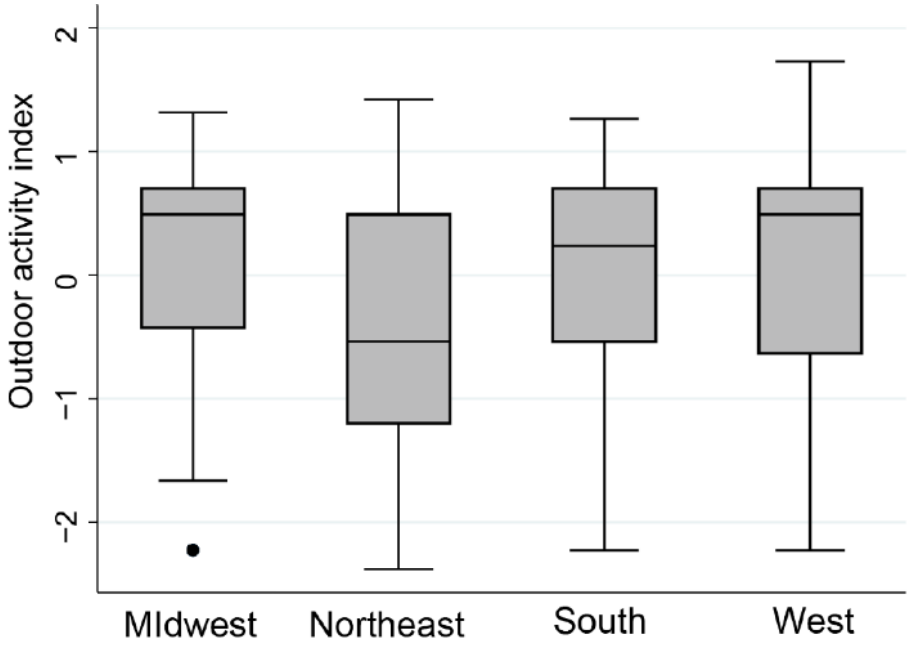


Table 3. Generalized linear mixed model for the number of users per county versus the county population size and the Lyme disease status of the county based on Lyme disease incidence and percent change of cases from 2013 to 2017, adjusting for the regions' random effects. We used a negative binomial model, and the effect of each independent variable is expressed as incidence rate ratios.

Variables[a]	Incidence rate ratio	95% CI	*P* value
County population size (per 100,000)	1.3	1.2-1.4	<.001[b]
County Lyme disease status (2013-2017)			
No Lyme disease cases reported	1	__[c]	—
Low incidence—no change	0.8	0.4-1.7	.60
Low incidence—greater than 1-fold higher increase	1.8	1.1-3.2	.03[d]
High incidence—no change	4.2	2.1-8.1	<.001[b]
High incidence—greater than 1-fold higher increase	3.5	1.8-7.2	<.001[b]

[a]Region (random effect) coefficient=0.6 (95% CI 0.1-4.9); Log-likelihood ratio test, $P<.001$.

[b]$P<.001$.

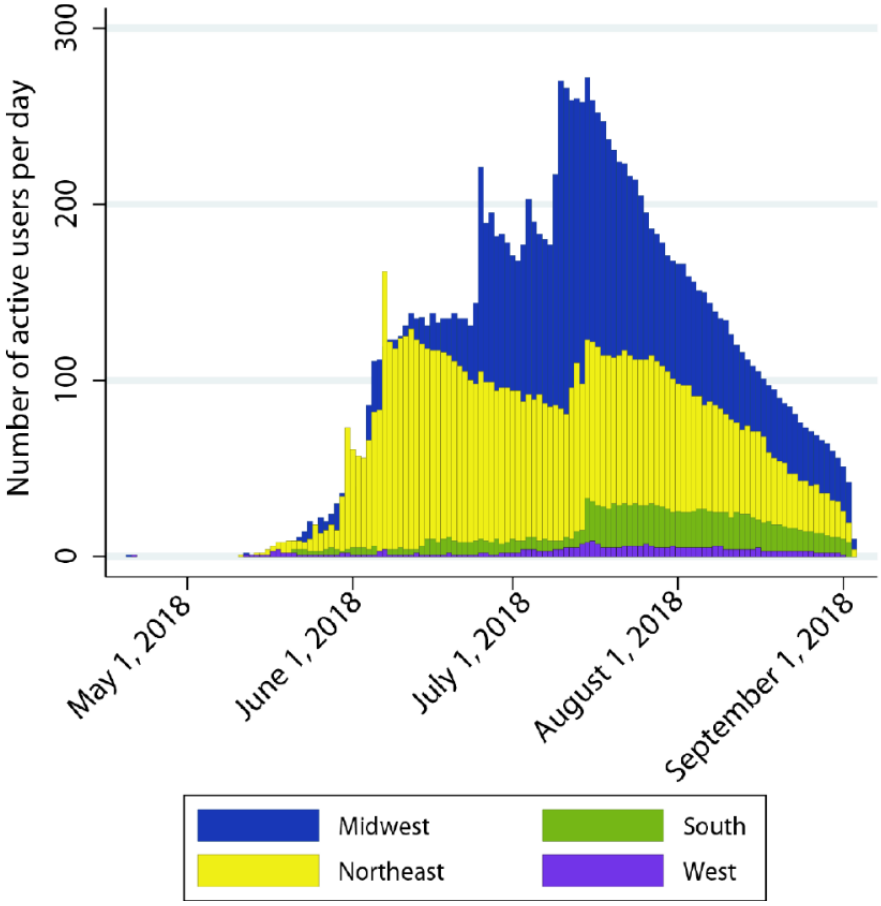
[c]Not applicable (reference category).

[d]$.001 \leq P \leq .05$.

Longitudinal App Use

After completing the enrollment survey, 49.25% (723/1468) of the users continued to interact with The Tick App. Users interacted with the app for a median of 25 days (IQR 12-49), 17.5% (126/723) interacted for up to 1 week, 19.0% (138/723) interacted between 1 week and 15 days, and 63.5% (459/723) interacted for 15 days or more. The median time that users interacted with the app per day was 1.9 min (IQR 0.7-4.7), and the median number of total users per day was 251 (IQR 109-291), peaking in June and mid-July and declining steadily thereafter in both the Midwest and Northeast (Figure 6), consistent with a decline in nymphal activity.

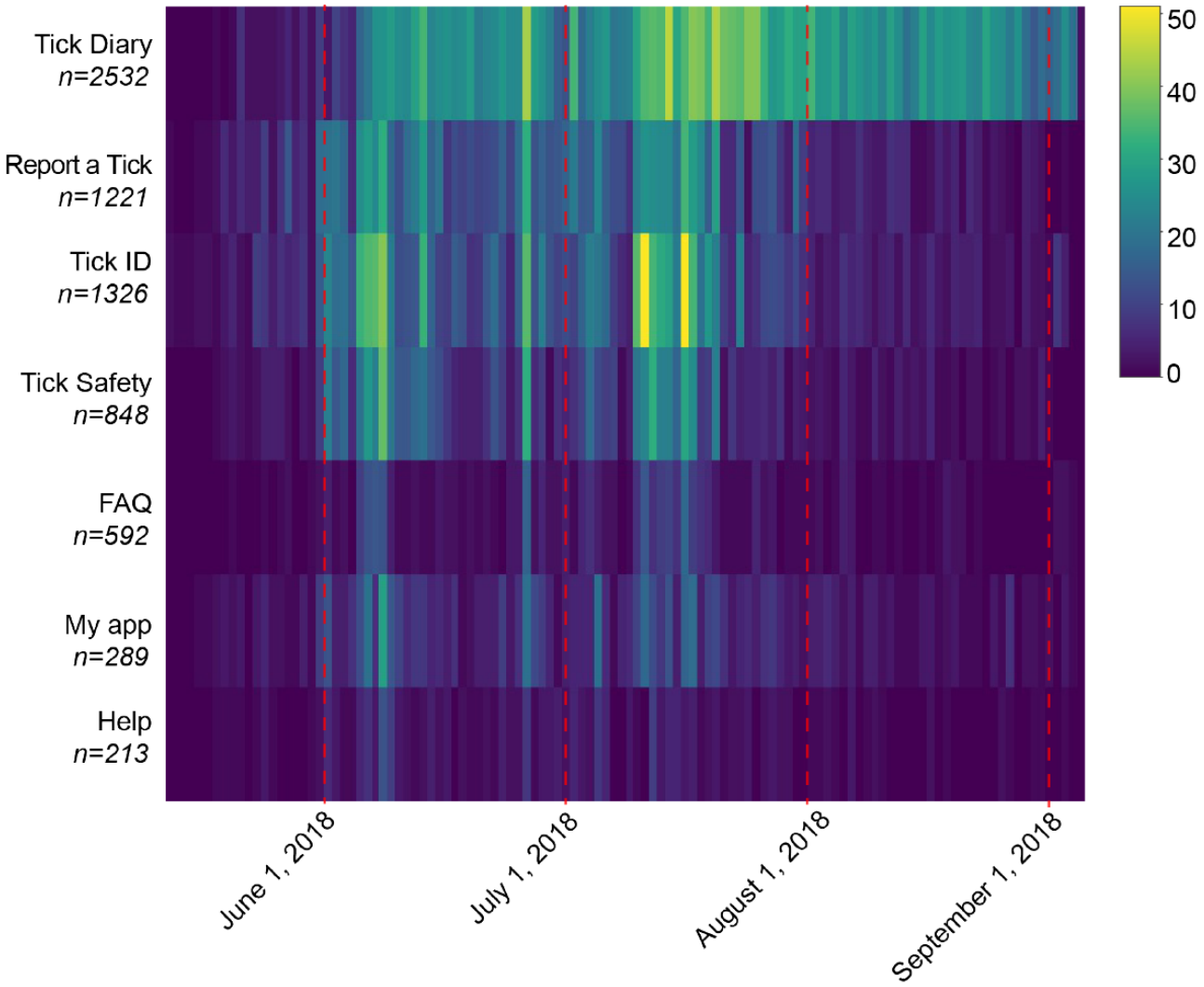
Figure 6. Number of active users per day between May and September 2018, by region.



The demographic profile of recurring users compared with nonrecurring users did not vary with gender or age, but users were more likely to use the app after completing the baseline survey if they worked or volunteered outdoors and if they did outdoor recreational activities frequently, although no effect was observed if they did frequent peridomestic activities (Multimedia Appendix 10). A higher proportion of recurring users reported having had a previous tick bite compared with nonrecurring users (268/721, 37.2% vs n=191/742, 25.7%, χ^2_1=22.1; *P*<.001), but no differences were observed in the proportion of self-reported previous diagnosis of a tick-borne disease (95/722, 13.2% vs 78/742, 10.5%, χ^2_1=2.4; *P*=.12). Similarly, the number of follow-up days per user increased only with the outdoor activity index (IRR 1.2, 95% CI 1.1-1.4; *P*<.01), after adjusting for age, gender, and the region, indicating that users were more likely to use the app if they did outdoor activities frequently.

On the basis of the unique screen views every 5 min (n=7021), the *Tick Diary* was the most frequently used feature in The Tick App, followed by *Report a Tick* and *Tick ID* (Figure 7). Educational material (*Tick safety*, including FAQ) was less popular among users, and app-related information (*Help* and *My App*) was the least used feature (Figure 7). The *Tick Diary* use was sustained during the study period, peaking in mid-July, following a peak in active users (Figures 6 and 7). *Report a Tick*, *Tick ID*, and *Tick Safety* had a more variable longitudinal use and were most frequently used in early June and mid-July (Figure 7).

Figure 7. Heatmap of the independent screen views (every 5 min) per functionality, for all users between May and September 2018. For each day and functionality, the lighter color (see color scale) indicates a higher number of independent screen views. The dashed red lines indicate the first day of each month for reference.



At least one *Tick Diary* was completed by 50.8% (367/723) of the recurring users, and one-fourth of the total users enrolled in the app. The proportion of recurring users completing *Tick Diaries* for at least 1 week was 32.2% (118/367), whereas 21.8% (80/367) of recurring users completed at least 15. In total, 367 *Tick Diaries* were completed during the study period, with a median number of 2 submitted *Tick Diaries* per user (IQR 1-11). The median time to complete and submit each *Tick Diary* was 34 seconds (IQR 22-63). The users reported a tick encounter on themselves in 24.5% (90/367) of the *Tick Diaries* submitted, and a total of 149 tick encounters, which resulted in 12 tick encounters per 1000 person-days. In addition, tick encounters on pets were reported in 17.0% (46/270) of *Tick Diaries*, whereas tick encounters on a member of the family were reported in 13.4% (43/320) of *Tick Diaries*.

Similar to the *Tick Diary*, approximately half of the recurring users (381/723, 52.7%) and one-fourth (381/1468, 26.0%) of the total number of enrolled users submitted at least one *Report a Tick* survey. However, they were not always the same recurring users completing both surveys; only 42.5% (162/381) of the users who completed at least one *Tick Diary* also submitted a tick report using the *Report a Tick* functionality. In total, 650 tick reports were submitted, with a median number of 1 report per user (IQR 1-2). The median time to complete *Report a Tick* was 1.8 min (IQR 1.1-3.3). Although the number of tick reports per user increased significantly with the outdoor activity index and women submitted more *Tick Diaries* than men, the number of *Tick Diaries* completed did not vary significantly with the frequency of outdoor activities or gender (Table 4). The number of both *Tick Diaries* and tick reports was significantly higher in older age groups (Table 4). Regional differences in the number of *Tick Diaries* per user were observed; users living the Midwest submitted more *Tick Diaries* than completed users in the Northeast, although these regional differences were not observed in the number of tick reports (Table 4).

Table 4. Generalized linear model for the number of Tick Diaries and tick reports submitted per user versus demographic variables, location (region), and frequency of outdoor activities (outdoor activity index). We used a negative binomial model, and the effect of each independent variable is expressed as incidence rate ratios (N=712).

Variables	Number of *Tick Diaries*			Number of tick reports		
	IRR[a]	95% CI	*P* value	IRR[a]	95% CI	*P* value
Gender						
Female	1	__[b]	—	1	—	—
Male	0.9	0.6-1.3	0.66	0.7	0.5-0.9	.01[c]
Age (years)						
18-25	1	—	—	1	—	—
25-35	1.1	0.5-2.5	.86	1.8	0.9-3.6	.07
35-45	1.2	0.5-2.7	.75	1.4	0.7-2.7	.32
45-55	1.3	0.6-3.1	.49	1.3	0.6-3.1	.48
55-65	3.4	1.5-7.6	<.001[d]	1.6	0.8-3.1	.15
≥65	3.8	1.6-9.2	<.001[d]	2.3	1.1-4.5	.02[c]
Region						
Midwest	1	—	—	1	—	—
Northeast	0.4	0.3-0.7	<.001[d]	1	0.7-1.3	.98
South	0.5	0.2-1.1	.10	1	0.6-1.6	.96
West	0.3	0.1-1.3	.10	1.2	0.4-3.2	.73
Outdoor activity index	1.1	0.9-1.4	.41	1.5	1.3-1.8	<.001[d]

[a]IRR: incidence rate ratio.

[b]Not applicable (reference category).

[c].001≤*P*≤.05.

[d]*P*<.001.

Feasibility Evaluation

We identified 4 major themes from the feedback provided by users at the end of the study period: (1) effective communication, (2) content, (3) operability, and (4) incentives. Effective communication referred to our ability to effectively communicate the goals of the study and the intended use of the app. We found that some users were confused about completing the *Tick Diaries* and thought they only had to complete one when they found a tick, although we requested 15 consecutive *Tick Diaries* regardless of tick encounter. The main reasons why we failed to communicate effectively appeared to be the name of the functionality (ie, *Tick Diaries* included the word Tick and that was misleading) and the insufficient explanation before accessing the homepage of the app. Although this information was included in the Help functionality, it was rarely accessed by users. Users participating in the focus groups agreed that the educational content was complete, but some mentioned some difficulties in navigating the app to find some of the materials they were most interested in (eg, how to remove a tick). They mentioned greater interest in practical resources than in general knowledge about ticks and tick-borne diseases. Regarding operability, users mentioned that it was easy to navigate and access, although some issues about the content organization were mentioned. In general, the operability of The Tick App matched the skills of the users and did not require special training. Finally, as in the preoperationalization feedback, users mentioned that they would like to access more local data regarding the risk of tick encounters and mentioned mapping reported ticks in their neighborhood as a highly desirable functionality.

Discussion

We used an iterative mixed-methods approach, in which findings from each step informed subsequent steps, to develop and evaluate the implementation of The Tick App, the first theory- and evidence-based smartphone app available in the United States to focus on ticks and tick-borne diseases. The Tick App users were equally represented across genders and evenly distributed across age groups; most users owned a pet, did frequent outdoor activities (recreational or peridomestic), and lived in the Midwest and Northeast regions in the United States, more specifically in Wisconsin, southern New York, and New Jersey. The number of users was higher in high-incidence counties for Lyme disease and in those counties with cases recently increasing. Half of the users were recurring users who participated in outdoor activities more frequently.

Within the wide range of citizen science projects, The Tick App can be classified as *contributory* because the public is mostly

involved in data collection [42]. Owing to the public's voluntary role, it is imperative to appeal to the users' needs and evaluate the uptake and longitudinal use of the app to validate its use as a research app. In this study, qualitative analyses were used to test and adapt multiple versions of the app to maximize users' experience. In line with van Velsen et al [29] who identified the requirements for a smartphone app to engage with the public in the prevention of tick bites in the Netherlands, The Tick App included a functionality to allow users to report ticks, videos with information on how to remove an attached tick, and alerts to remind people to check for ticks at the end of the day. The latter was also coupled with an alert to complete the daily activity survey, in line with the EMA method used in other human behavioral studies [43]. Quantitative analyses of the data actively submitted by the users and passively collected by the app (ie, usage data) were used to assess potential biases in the data collected and evaluate the factors associated with the uptake and longitudinal use of the app.

The Tick App is unique in its design to provide a bidirectional flow of tick-related information (from the user to the researcher and vice versa). Users were able to receive tick identification services and educational information. In turn, our research team was able to combine tick reports with epidemiological data related to human behavior to better understand the risk of tick exposure. Other tick-related smartphone apps have been implemented in Europe, either by research teams (Tekenbeet and Tekenscanner in the Netherlands) [10,11] or public health agencies (Zecke in Switzerland, Signalement TIQUE in France, and TekenNet/TiquesNet in Belgium) [44-46]. These apps mainly focus on educational materials and tick surveillance from user reports but do not describe specific research objectives related to the behavioral risk factors of tick exposure. In addition, these apps are intended to be used at the country level [29]. In the United States, commercial apps focus on automated tick identification (What's My Tick and Detectick) [47,48] and mapping of tick reports by users (TickTracker) [49]. The latter creates heatmaps (ie, density maps) of tick reports but without a validation procedure to ensure the reliability of the data, that is, the species identification and whether the specimen is a tick. With the exception of Tekenbeet [10,29], no publicly available information is provided for the other apps, which describes the methods and considerations taken in the design of the app or in the uptake and usage by the general public.

The geographic distribution of The Tick App users indicated that we were targeting the population residing in regions with the greatest risk of tick exposure to *I scapularis* ticks. Although The Tick App was available nationwide, the enrollment of users was higher in the Northeast and Upper Midwest regions in the United States, where the majority of the Lyme disease cases were reported [12]. Within these regions, uptake of the app was highest in those counties with high Lyme disease incidence and with recent increases in the number of reported cases. Nonetheless, the uptake of the app further coincided with those areas where we conducted active and passive recruitment of participants (Wisconsin and New York State), indicating that the population exposed to a higher risk of tick encounters is not necessarily actively searching for a smartphone app to help them manage the risk but is receptive to its use. We also observed a higher proportion of iOS users compared with Android users, although the market share trend was the opposite as of 2019 [50], which could indicate differences in socioeconomic status, age, and use pattern [51].

According to the COM-B system, a human behavioral model that emerged from a systematic review of behavior change theories [52,53], enrollment in the app may depend on the opportunities provided by the app (functionalities that appeal to the users' needs), capability (confidence in using the app or check for ticks), and motivation (risk perception, sense of empowerment, or expectation of reward). When analyzing the user profile, the majority of users did report frequent outdoor recreational and peridomestic activities, and half of the users also reported outdoors occupations, which have previously been associated with the risk of tick encounters [13,14,16,22]. In this study, because of the high correlation between the variables related to outdoor exposure, we created an overall index of frequent outdoor activities, which was also found to be associated with self-reported tick encounters. Self-reported previous tick exposure of the app users was slightly higher (31.4%) than those reported in the national HealthStyles surveys for the Upper Midwest and Northeast (22.6%-29.8%), in which respondents were randomly recruited from a large, nationally representative panel of adults aged 18 years or older [20]. However, when comparing self-reported diagnosis with previous tick-borne disease diagnosis, the proportion of users in the app (11.8%) almost doubled that of respondents in the HealthStyles surveys (2.0%-6.5%) [20], and the proportion of users reporting Lyme disease cases in the previous year was higher compared with the general population in the counties where the users lived. These results indicate that The Tick App users are biased toward those with a previous self-reported diagnosis with a tick-borne disease rather than toward those with previous self-reported tick exposure. However, self-reported Lyme disease diagnosis might be higher than reported and estimated cases [54], and previous tick encounters will depend on the user's frequency of checking for ticks after being outdoors and their ability to detect and identify a tick. Nonetheless, we achieved a broader representation of the general population compared with the Tekenbeet app in the Netherlands, where they found that 90.9% of their users reported having had a previous tick encounter and 56.4% reported previous Lyme disease diagnosis. In The Tick App, we observed no other strong demographic bias when analyzing the users' profile, whereas Tekenbeet was slightly more biased toward females (56.8%).

Analysis of the longitudinal use of The Tick App indicated that its use was seasonal and coincided with the trend observed for google searches for *Lyme disease* and *tick bites* [55]. In the case of The Tick App, 49.2% of initial users continued using the app beyond enrollment compared with 67.5% of Tekenbeet users in the Netherlands, which has very similar features (tick reports and educational material) but does not include any surveys to collect behavioral information or other risk factors. Continued use of The Tick App was also related to the frequency of outdoor activities. Motivation to enroll and initial participation in citizen science projects have been found to be related to motivations pertaining to one's own welfare (*egoism*), whereas increasing welfare of others and of the group seemed to play an important

role (*altruism* and *collectivism*, respectively) for continuing participation [42]. *Altruism* and *collectivism* seemed to be important motivational drivers, given that the *Tick Diary* was the most frequently used functionality and it does not directly reward the user; however, the mean number of *Tick Diaries* submitted per person did not achieve our research goals (15 complete *Tick Diaries* per user). From the users' feedback at the end of the study period, we identified 2 main barriers: (1) the objective was not clear to participants and (2) we needed new motivational strategies to sustain engagement of the initial participants. The *egoism* motivational category for continued use of The Tick App was still of considerable importance: the *Report a Tick* and *Tick ID* features were among the most frequently used as well. Although the former feature is also intended for our research goals, the associated online picture submission and tick identification services provided by the research group could be the primary motivation for the user. The user profile differed between those engaged with the research via *Tick Diaries* and those engaging via *Report a Tick*. Users completing *Tick Diaries* were older (aged 55 years and older) and might have more time to participate, whereas users completing tick reports engaged more often in outdoor activities. The challenge is to respond to the different motivational factors at different points of participation to keep users engaged and achieve a broader representation of the population at risk of tick bites.

Additional incentives for The Tick App could help increase enrollment and engagement but need to be carefully selected because they can introduce potentially unintended consequences [56]. One functionality that users mentioned that they would like in the app was a map showing the ticks that have been reported in their area. The TickTracker app allows users to map the ticks that have been reported through the app but lacks scientific quality control of the reports. There are 2 main issues with mapping self-reported data if the users are going to rely on this information for decision making: (1) the number of ticks reported will depend on the number of users in the area and (2) validation of tick reports (ie, confirming that specimens are ticks and determining which species is being reported) currently requires considerable time and human resources. Quality maps of tick distribution at a small scale (ie, county) would require a large number of validated tick reports provided by a large number of users continuously using the app. Mapping incomplete data from areas with few users and few reports could result in a false sense of safety. On the other hand, mapping unvalidated reports could result in a false sense of risk and create unnecessary anxiety in the user. An alternative to near real-time mapping of tick reports is providing an indicator of tick activity based on the location and seasonality, a strategy that was implemented in the Tekenbeet app at a national level. We are planning on incorporating this feature at a county level in a future iteration of The Tick App (The Tick App 2.0).

Tick identification seemed to be one of the main incentives for using the app, although picture submission was done externally to the app. In The Tick App 2.0, users will be able to submit a photo of the tick within the app to reduce the effort from the user endpoint. Automated tick identification built into the app would greatly reduce the resources invested in tick identification. However, the apps offering this functionality are either unclear about the validation procedures of the classification algorithm and certainty of tick species identification from a photo (What's My Tick) or do not offer tick identification to the species level (Detectick). The incentives to complete the *Tick Diaries* (the research aspect of the app) may require other types of incentives [43]. Gamification or the use of game design elements (badges, leaderboards, rewards, and avatars) can help maintain user engagement by "harnessing the desire for competition and the goal-driven aspects of human nature" [56,57]. In The Tick App 2.0, we will incorporate badges that can be earned as daily surveys are completed and a progress bar so that the users can track their progress toward completing a 7-day streak, a 15-day streak, and a 21-day streak. These streaks were established based on the results from the analysis of the longitudinal usage: approximately 30% of returning users completed at least 7 *Tick Diaries* and approximately 20% completed at least 15 *Tick Diaries*. The minimum of 7 days will allow for each day of the week to be represented and allow us to study daily fluctuations in activity [43]. We also changed the name to *Daily Log* because *Tick Diary* was understood by some users as only meant to be completed when a tick was found, and a clearer explanation of the objectives was also included in the app. Finally, monetary incentives can also be considered, particularly to increase representation among less motivated groups [23,56]. To achieve a broader representation of the daily activities and the risk of tick encounters in the general population, monetary incentives could be useful at a manageable scale and for a specific period and would increase the external validity of our results. A downside of monetary incentives is that they might result in data fabrication for monetary gain [56]; however, Bell et al [58] found that offering incentives to complete regular surveys through an app did not encourage false responses when the same reward was offered regardless of the answer, although 1 response required less information (and thus, less effort from the user).

A common goal for many mHealth studies is the "identification of behaviors associated with health outcomes so that behavior change interventions may be designed and implemented on a large scale" [59]. The analysis of the enrollment and use of The Tick App helped us identify the successful aspects of the app as well as its limitations and potential biases that could limit the extrapolation of the results derived from the data collected. The Tick App also offers the opportunity to explore interventions oriented to reduce the risk of tick-borne diseases by increasing self-awareness and encouraging the use of protective measures [10,23,60]. Understanding who, how, and when people are using The Tick App would help us tailor the content of an intervention to achieve a greater effect. Finally, there is a need to continually evaluate and revise the app based on what users are willing to do and what they can expect in return, while meeting the data requirements for the research on the behavioral risk factors of human-tick encounters.

Acknowledgments

The authors thank the NYC Department of Health and Mental Hygiene and Eau Claire City Health Department for advice and support; Paula Fernandez Begne, Cyrus Hadavi, Angel Weng, Vincent de Souza, Lara Ferreyra, and Just Mitchell Kirsch for field support and active recruitment of participants; and Mary Hayden for advice and leading the initial focus groups on Staten Island. The authors are also very grateful to all the users for their participation in The Tick App and focus groups. This publication was supported by Cooperative Agreements #U01 CK000505 and #U01 CK000509-01, funded by CDC. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC or the Department of Health and Human Services. The funders had no role in the study design, analysis or interpretation of data, or writing of the report.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Phase 1 and 2: Pre-prototype and prototype design.
[PDF File (Adobe PDF File), 68 KB - mhealth_v7i10e14769_app1.pdf]

Multimedia Appendix 2
Consolidated Standards of Reporting Trials electronic health checklist.
[PDF File (Adobe PDF File), 2590 KB - mhealth_v7i10e14769_app2.pdf]

Multimedia Appendix 3
Checklist for Reporting Results of Internet E-Surveys.
[PDF File (Adobe PDF File), 123 KB - mhealth_v7i10e14769_app3.pdf]

Multimedia Appendix 4
Technical appendix of the backend components of The Tick App.
[PDF File (Adobe PDF File), 67 KB - mhealth_v7i10e14769_app4.pdf]

Multimedia Appendix 5
Informed consent and recruitment material.
[PDF File (Adobe PDF File), 1866 KB - mhealth_v7i10e14769_app5.pdf]

Multimedia Appendix 6
Guiding questions for focus groups conducted prior and after The Tick App implementation.
[PDF File (Adobe PDF File), 82 KB - mhealth_v7i10e14769_app6.pdf]

Multimedia Appendix 7
Enrollment of new users per day between May and September 2018 and the cumulative number of users in the period. The vertical dashed lines indicate media coverage of The Tick App.
[PDF File (Adobe PDF File), 95 KB - mhealth_v7i10e14769_app7.pdf]

Multimedia Appendix 8
Age and gender distribution of users of The Tick App.
[PDF File (Adobe PDF File), 291 KB - mhealth_v7i10e14769_app8.pdf]

Multimedia Appendix 9
Results of the multiple correspondence analysis (MCA), including frequent outdoor activities (peridomestic and recreational), having an outdoor job and owning a pet.
[PDF File (Adobe PDF File), 31 KB - mhealth_v7i10e14769_app9.pdf]

Multimedia Appendix 10
The recurring users' profile including demographic variables, frequent outdoor activities (occupational, recreational and peridomestic), and owning a pet as reported in the enrollment survey.
[PDF File (Adobe PDF File), 111 KB - mhealth_v7i10e14769_app10.pdf]

References

1. Pew Research Center. 2019. Mobile Fact Sheet URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-04-01] [WebCite Cache ID 78SySSZ34]
2. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
3. Lefebvre C. Integrating cell phones and mobile technologies into public health practice: a social marketing perspective. Health Promot Pract 2009 Oct;10(4):490-494. [doi: 10.1177/1524839909342849] [Medline: 19809002]
4. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
5. Bell AR. Informing decisions in agent-based models — a mobile update. Environ Model Softw 2017 Jul;93:310-321. [doi: 10.1016/j.envsoft.2017.03.028]
6. Schmidt C, Phippard A, Olsen JM, Wirt K, Riviera A, Crawley A, et al. Kidenga: public engagement for detection and prevention of Aedes-borne viral diseases. Online J Public Health Inform 2017 May 2;9(1):e111. [doi: 10.5210/ojphi.v9i1.7694]
7. Eskenazi B, Quirós-Alcalá L, Lipsitt JM, Wu LD, Kruger P, Ntimbane T, et al. mSpray: a mobile phone technology to improve malaria control efforts and monitor human exposure to malaria control pesticides in Limpopo, South Africa. Environ Int 2014 Jul;68:219-226 [FREE Full text] [doi: 10.1016/j.envint.2014.03.003] [Medline: 24769412]
8. Reddy E, Kumar S, Rollings N, Chandra R. ArXiv. 2015. Mobile Application for Dengue Fever Monitoring and Tracking via GPS: Case Study for Fiji URL: https://arxiv.org/abs/1503.00814 [accessed 2015-04-01]
9. Meankaew P, Kaewkungwal J, Khamsiriwatchara A, Khunthong P, Singhasivanon P, Satimai W. Application of mobile-technology for disease and treatment monitoring of malaria in the 'Better Border Healthcare Programme'. Malar J 2010;9:237 [FREE Full text] [doi: 10.1186/1475-2875-9-237] [Medline: 20723223]
10. Antonise-Kamp L, Beaujean DJ, Crutzen R, van Steenbergen JE, Ruwaard D. Prevention of tick bites: an evaluation of a smartphone app. BMC Infect Dis 2017 Dec 4;17(1):744 [FREE Full text] [doi: 10.1186/s12879-017-2836-4] [Medline: 29202704]
11. Jongejan F, de Jong S, Voskuilen T, van den Heuvel L, Bouman R, Heesen H, et al. 'Tekenscanner': a novel smartphone application for companion animal owners and veterinarians to engage in tick and tick-borne pathogen surveillance in the Netherlands. Parasit Vectors 2019 Mar 26;12(1):116 [FREE Full text] [doi: 10.1186/s13071-019-3373-3] [Medline: 30909941]
12. Schwartz AM, Hinckley AF, Mead PS, Hook SA, Kugeler KJ. Surveillance for Lyme disease - United States, 2008-2015. MMWR Surveill Summ 2017 Nov 10;66(22):1-12 [FREE Full text] [doi: 10.15585/mmwr.ss6622a1] [Medline: 29120995]
13. Mead P, Hook S, Niesobecki S, Ray J, Meek J, Delorey M, et al. Risk factors for tick exposure in suburban settings in the Northeastern United States. Ticks Tick Borne Dis 2018 Feb;9(2):319-324. [doi: 10.1016/j.ttbdis.2017.11.006] [Medline: 29174449]
14. Finch C, Al-Damluji MS, Krause PJ, Niccolai L, Steeves T, O'Keefe CF, et al. Integrated assessment of behavioral and environmental risk factors for Lyme disease infection on Block Island, Rhode Island. PLoS One 2014;9(1):e84758 [FREE Full text] [doi: 10.1371/journal.pone.0084758] [Medline: 24416278]
15. Connally NP, Durante AJ, Yousey-Hindes KM, Meek JI, Nelson RS, Heimer R. Peridomestic Lyme disease prevention: results of a population-based case-control study. Am J Prev Med 2009 Oct;37(3):201-206. [doi: 10.1016/j.amepre.2009.04.026] [Medline: 19595558]
16. Berry K, Bayham J, Meyer SR, Fenichel EP. The allocation of time and risk of Lyme: a case of ecosystem service income and substitution effects. Environ Resour Econ (Dordr) 2018 Jul;70(3):631-650 [FREE Full text] [doi: 10.1007/s10640-017-0142-7] [Medline: 30166775]
17. Kitron UD, Kazmierczak JJ. Spatial analysis of the distribution of Lyme disease in Wisconsin. Am J Epidemiol 1997 Mar 15;145(6):558-566. [doi: 10.1093/oxfordjournals.aje.a009145] [Medline: 9063347]
18. Hahn MB, Bjork JK, Neitzel DF, Dorr FM, Whitemarsh T, Boegler KA, et al. Evaluating acarological risk for exposure to Ixodes scapularis and Ixodes scapularis-borne pathogens in recreational and residential settings in Washington County, Minnesota. Ticks Tick Borne Dis 2018 Feb;9(2):340-348. [doi: 10.1016/j.ttbdis.2017.11.010] [Medline: 29195857]
19. Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, et al. Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking Ixodes scapularis nymphs in the Eastern United States. Am J Trop Med Hyg 2012 Jul;86(6):1062-1071 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0630] [Medline: 22665620]
20. Hook SA, Nelson CA, Mead PS. US public's experience with ticks and tick-borne diseases: results from national HealthStyles surveys. Ticks Tick Borne Dis 2015 Jul;6(4):483-488. [doi: 10.1016/j.ttbdis.2015.03.017] [Medline: 25887156]
21. Hinckley AF, Meek JI, Ray JA, Niesobecki SA, Connally NP, Feldman KA, et al. Effectiveness of residential acaricides to prevent Lyme and other tick-borne diseases in humans. J Infect Dis 2016 Jul 15;214(2):182-188. [doi: 10.1093/infdis/jiv775] [Medline: 26740276]
22. Lane RS, Steinlein DB, Mun J. Human behaviors elevating exposure to Ixodes pacificus (Acari: Ixodidae) nymphs and their associated bacterial zoonotic agents in a hardwood forest. J Med Entomol 2004 Mar;41(2):239-248. [doi: 10.1603/0022-2585-41.2.239] [Medline: 15061284]

23. Runyan JD, Steenbergh TA, Bainbridge C, Daugherty DA, Oke L, Fry BN. A smartphone ecological momentary assessment/intervention. PLoS One 2013;8(8):e71325 [FREE Full text] [doi: 10.1371/journal.pone.0071325] [Medline: 23977016]
24. Fernandez MP, Bron GM, Tsao JI, Bartholomay LC, Paskewitz SM, Diuk-Wasser MA. The Tick App. URL: https://thetickapp.org/ [accessed 2019-04-26] [WebCite Cache ID 77vQOLPyS]
25. Shiffman S. Conceptualizing analyses of ecological momentary assessment data. Nicotine Tob Res 2014 May;16(Suppl 2):S76-S87 [FREE Full text] [doi: 10.1093/ntr/ntt195] [Medline: 24323571]
26. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
27. Watkins KL, Regan SD, Nguyen N, Businelle MS, Kendzor DE, Lam C, et al. Advancing cessation research by integrating EMA and geospatial methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. Nicotine Tob Res 2014 May;16(Suppl 2):S93-101 [FREE Full text] [doi: 10.1093/ntr/ntt135] [Medline: 24057995]
28. World Health Organization. 2016. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment URL: https://apps.who.int/iris/bitstream/handle/10665/252183/9789241511766-eng.pdf [accessed 2019-12-08]
29. van Velsen L, Beaujean DJ, Wentzel J, van Steenbergen JE, van Gemert-Pijnen J. Developing requirements for a mobile app to support citizens in dealing with ticks and tick bites via end-user profiling. Health Informatics J 2015 Mar;21(1):24-35. [doi: 10.1177/1460458213496419] [Medline: 24282189]
30. Greenacre M. La Práctica Del análisis De Correspondencias. Bilbao, Spain: Rubes Editorial; 2008.
31. Sourial N, Wolfson C, Zhu B, Quail J, Fletcher J, Karunananthan S, et al. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. J Clin Epidemiol 2010 Jul;63(6):638-646 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.08.008] [Medline: 19896800]
32. Asselin LM, Anh VT. Multidimensional poverty measurement with multiple correspondence analysis. In: Kakwani N, Silber J, editors. Quantitative Approaches to Multidimensional Poverty Measurement. London: Palgrave Macmillan; 2008:80-103.
33. United States Census Bureau. 2019. Census Regions and Divisions of the United States URL: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [accessed 2019-05-20] [WebCite Cache ID 78W0QM9My]
34. Center for Disease Control and Prevention. 2018. Lyme Disease Surveillance and Available Data: Background URL: https://www.cdc.gov/lyme/stats/survfaq.html [accessed 2019-04-29] [WebCite Cache ID 780DKGhNJ]
35. Stockwell EG, Shryock HS, Siegel JS. The Methods and Materials of Demography. Second Edition. London: Emerald Group Publishers; 2004.
36. Quinn GP, Keough MJ. Experimental Design and Data Analysis for Biologists. Cambridge: Cambridge University Press; 2012.
37. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press; 2007.
38. Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. Mixed Effects Models and Extensions in Ecology With R. New York: Springer; 2009.
39. Zeileis A, Kleiber C, Jackman S. Regression models for count data in R. J Stat Soft 2008;27(8):1-25. [doi: 10.18637/jss.v027.i08]
40. Stata: software for statistics and data science. 2017. URL: https://www.stata.com/ [accessed 2019-04-27]
41. The R Project for statistical computing. 2018. URL: https://www.r-project.org/ [accessed 2019-04-27]
42. Land-Zandstra AM, Devilee JL, Snik F, Buurmeijer F, van den Broek JM. Citizen science on a smartphone: participants' motivations and learning. Public Underst Sci 2016 Jan;25(1):45-60. [doi: 10.1177/0963662515602406] [Medline: 26346340]
43. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854. [doi: 10.1177/1745691616650285] [Medline: 27899727]
44. National Reference Centre for Tick-borne Diseases (NRZK). Zeke. Use of the Interactive App URL: https://zecke-tique-tick.ch/en/app-tick/ [accessed 2019-04-27] [WebCite Cache ID 77w3Hd6lf]
45. Institut National de la Recherche Agronomique (INRA), Ministère des Solidarités et de la Santé. Signalement TIQUE URL: http://ephytia.inra.fr/fr/Products/view/159/Signalement_TIQUE [accessed 2019-04-01] [WebCite Cache ID 77w38NWUx]
46. Scientific Institute of Public Health (Sciensano). TicksNet – TekenNet – TiquesNet URL: https://epistat.wiv-isp.be/ticks/ [accessed 2019-04-27] [WebCite Cache ID 77w3TRBlY]
47. Bilodeau R. App Store - Apple. What's My Tick URL: https://itunes.apple.com/app/id1431587761 [accessed 2019-04-01] [WebCite Cache ID 77w2RwPQ3]
48. Guertin A. Google Play. Detectick URL: https://play.google.com/store/apps/details?id=detectick.com.detectick&hl=en_US [accessed 2019-04-01] [WebCite Cache ID 77w2RwPQ3]
49. LivLyme Foundation. TickTracker App. URL: https://ticktracker.com/ [accessed 2019-04-27] [WebCite Cache ID 77w3Z0cRa]

50. Holst A. Statista. 2019. Subscriber share held by smartphone operating systems in the United States from 2012 to 2019 URL: https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/ [accessed 2019-08-01]
51. Ubhi HK, Kotz D, Michie S, van Schayck OC, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171 [FREE Full text] [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]
52. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
53. Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, et al. Developing and evaluating JIApp: acceptability and usability of a smartphone app system to improve self-management in young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2017 Aug 15;5(8):e121 [FREE Full text] [doi: 10.2196/mhealth.7229] [Medline: 28811270]
54. Kuehn BM. CDC estimates 300 000 US cases of Lyme disease annually. JAMA 2013 Sep 18;310(11):1110. [doi: 10.1001/jama.2013.278331] [Medline: 24045727]
55. Seifter A, Schwarzwalder A, Geis K, Aucott J. The utility of 'Google Trends' for epidemiological research: Lyme disease as an example. Geospat Health 2010 May;4(2):135-137. [doi: 10.4081/gh.2010.195] [Medline: 20503183]
56. Hines D, Sibbald SL. Citizen science: exploring its application as a tool for prodromic surveillance of vector-borne disease. Can Commun Dis Rep 2015 Mar 5;41(3):63-67 [FREE Full text] [doi: 10.14745/ccdr.v41i03a04] [Medline: 29769934]
57. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
58. Bell AR, Ward PS, Killilea ME, Tamal ME. Real-time social data collection in rural Bangladesh via a 'microtasks for micropayments' platform on Android smartphones. PLoS One 2016;11(11):e0165924 [FREE Full text] [doi: 10.1371/journal.pone.0165924] [Medline: 27832116]
59. Lathia N, Pejovic V, Rachuri KK, Mascolo C, Musolesi M, Rentfrow PJ. Smartphones for large-scale behavior change interventions. IEEE Pervasive Comput 2013 Jul;12(3):66-73. [doi: 10.1109/MPRV.2013.56]
60. Kramer JN, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]

Abbreviations

CDC: Centers for Disease Control and Prevention
EMA: ecological momentary assessment
GPS: global positioning system
IQR: interquartile range
IRB: Institutional Review Board
IRR: incidence rate ratio
MCA: multiple correspondence analysis
mHealth: mobile health
OR: odds ratio

Edited by G Eysenbach; submitted 20.05.19; peer-reviewed by S Butler, S Lee; comments to author 18.06.19; revised version received 14.08.19; accepted 17.08.19; published 24.10.19.

Please cite as:
Fernandez MP, Bron GM, Kache PA, Larson SR, Maus A, Gustafson Jr D, Tsao JI, Bartholomay LC, Paskewitz SM, Diuk-Wasser MA
Usability and Feasibility of a Smartphone App to Assess Human Behavioral Factors Associated with Tick Exposure (The Tick App): Quantitative and Qualitative Study
JMIR Mhealth Uhealth 2019;7(10):e14769
URL: http://mhealth.jmir.org/2019/10/e14769/
doi:10.2196/14769
PMID:31651409

Original Paper

A Vendor-Independent Mobile Health Monitoring Platform for Digital Health Studies: Development and Usability Study

Thijs Vandenberk[1,2], MSc, PhD; Valerie Storms[1,2], MSc, PhD; Dorien Lanssens[1,2,3], MSc, PhD; Hélène De Cannière[1,2,4], MSc; Christophe JP Smeets[1,2,4], MSc, PhD; Inge M Thijs[1,2,3], MSc, PhD; Tooba Batool[5], MSc; Yves Vanrompay[5], PhD; Pieter M Vandervoort[1,2,4], Prof MD; Lars Grieten[1,2], MSc, PhD

[1]Faculty of Medicine and Life Sciences, Hasselt University, Diepenbeek, Belgium

[2]Future Health Department, Hospital Oost-Limburg, Genk, Belgium

[3]Department of Obstetrics & Gynecology, Ziekenhuis Oost-Limburg, Genk, Belgium

[4]Department of Cardiology, Hospital Oost-Limburg, Genk, Belgium

[5]Transportation Research Institute (IMOB), Hasselt University, Hasselt, Belgium

Corresponding Author:
Thijs Vandenberk, MSc, PhD
Faculty of Medicine and Life Sciences
Hasselt University
Agorlaan
Diepenbeek, 3590
Belgium
Phone: 32 473900054
Email: thijsvandenberk@hotmail.com

Abstract

Background: Medical smartphone apps and mobile health devices are rapidly entering mainstream use because of the rising number of smartphone users. Consequently, a large amount of consumer-generated data is being collected. Technological advances in innovative sensory systems have enabled data connectivity and aggregation to become cornerstones in developing workable solutions for remote monitoring systems in clinical practice. However, few systems are currently available to handle such data, especially for clinical use.

Objective: The aim of this study was to develop and implement the digital health research platform for mobile health (DHARMA) that combines data saved in different formats from a variety of sources into a single integrated digital platform suitable for mobile remote monitoring studies.

Methods: DHARMA comprises a smartphone app, a Web-based platform, and custom middleware and has been developed to collect, store, process, and visualize data from different vendor-specific sensors. The middleware is a component-based system with independent building blocks for user authentication, study and patient administration, data handling, questionnaire management, patient files, and reporting.

Results: A prototype version of the research platform has been tested and deployed in multiple clinical studies. In this study, we used the platform for the follow-up of pregnant women at risk of developing pre-eclampsia. The patients' blood pressure, weight, and activity were semi-automatically captured at home using different devices. DHARMA automatically collected and stored data from each source and enabled data processing for the end users in terms of study-specific parameters, thresholds, and visualization.

Conclusions: The increasing use of mobile health apps and connected medical devices is leading to a large amount of data for collection. There has been limited investment in handling and aggregating data from different sources for use in academic and clinical research focusing on remote monitoring studies. In this study, we created a modular mobile health research platform to collect and integrate data from a variety of third-party devices in several patient populations. The functionality of the platform was demonstrated in a real-life setting among women with high-risk pregnancies.

(JMIR Mhealth Uhealth 2019;7(10):e12586) doi:10.2196/12586

KEYWORDS

information science; patient care management; mobile health; telemonitoring; monitoring, ambulatory

Introduction

Background

The increasing penetration of smartphones into consumer markets, as well as the growth in connected devices for health care, sport, and wellness, is leading to a dramatic increase in consumer-generated data [1]. Smartphones are becoming increasingly integrated into the global population. The number of smartphone users increased by nearly 1 billion people between 2014 (1.57 billion) and 2017 (2.32 billion), with an expected further increase to 2.87 billion users by 2020 [2].There are currently more than 259,000 health-related smartphone apps available in Web-based app stores; most of these apps were developed for Android (Google, CA, USA)and iOS (Apple Inc, CA, USA) [3]. The same technologies and concepts can also be applied to support health care. Remote monitoring of patients' vital signs and behaviors in their home environment could play an essential role in achieving a sustainable and high-quality health care system [4,5] and reduce health care costs [3,5,6] by offering near-continuous patient follow-up, better data management strategies, and better treatment adherence. Some examples of chronic diseases that could benefit from this type of platform include diabetes, heart failure, and cardiac arrhythmias [7-9]. The information obtained by remote monitoring provides an additional dimension to that possible with standard clinical care that is undertaken mainly by spot checks in the clinic. Continuous real-time patient tracking and processing of various parameters will influence the way health care practitioners deal with prevention, diagnostics, and disease management.

It is already possible to connect mobile health apps to a wide range of portable devices or sensors [10], such as those used to measure physiological parameters (eg, blood pressure, blood glucose, weight, and activity) [11]. An electrocardiogram (ECG) can be recorded via an app connected to a wearable ECG device [12]. In addition to the advantages for patients, doctors will also benefit from mobile health tools. Mobile apps can be used as an interface between the patient and the clinician via a communication channel [10,13], which could act as a substitute for some ambulatory visits [14]. As they enable patients to be monitored in their home environment, mobile health tools could prevent some hospitalizations or shorten the hospital stay [3,6] and have positive effects on medical outcomes, health care expenditure [6], and quality of life. These benefits would be due partly because the patient stays in their familiar environment and partly because continuous monitoring gives the opportunity for immediate intervention, if needed [10].

Digital Equipment

Remote monitoring patients at home requires a data communication system between the patient and the health care professional. Typically, the patient is equipped at home with a set of sensors that measure vital signs and a smartphone app that collects data on behavior, symptoms, and location. These data are collected on a cloud-based server that is accessed by the health care professional. Tools for decision support to assist the health care professional and semiautomatic feedback to patients are necessary to manage large patient populations.

Despite the significant technological advances in developing novel sensory systems and state-of-the-art devices, there has been limited investment in developing the infrastructure that is required to connect and handle the amount of information that these devices generate. In particular, there are limited tools available to handle this health information in terms of clinical applications [15]. With the rapid increase in novel tools and technologies, data connectivity and aggregation have become the cornerstones in developing workable solutions to manage patients in clinical practice and support scientific research in the provision of digital health [16]. However, the commercial remote monitoring technology market is highly fragmented because each vendor has developed their own data platform to record data from their associated sensors and communicate these using stand-alone software solutions or Web-based apps. Accordingly, it is impossible to aggregate data obtained from the sensors developed by different vendors. Clinical practice and academic research that rely on remote monitoring are limited by these closed, manufacturer-owned platforms. Therefore, a generic and open digital research platform for remote monitoring is needed to allow academic and clinical research into remote monitoring. The platform is required to overcome the problems of third-party device integration and the collection of various data feeds from patient populations.

Objectives

In this paper, we discuss the development of a generic and open digital research platform for remote monitoring that can be used to perform academic and clinical research into remote monitoring. This digital platform overcomes the problems of third-party device integration and collection of various data feeds from patient populations. The collected data can be analyzed using the platform and presented visually to the caregiver. We demonstrate the functionality of the platform in a real-life setting, among women with high-risk pregnancies.

Methods

Overview

The concept and design of our modular digital health research platform for mobile health (DHARMA) platform were focused primarily on clinical usability. Accordingly, health care professionals (doctors and nurses) participated in each stage of its conceptualization and development, from design and prototyping through to implementation and user trials.

The DHARMA platform was developed in close collaboration with all stakeholders: doctors, professional caregivers, researchers, and experts in mobile health. As its primary aim is to support clinical studies, the platform was built taking into account the requirements of specific studies but always with general applicability and genericity in mind. We ensured short test cycles by using the agile development methodology

(Scrum). An appropriate toolset was put into place for managing user stories (Jira), documenting design and implementation decisions (Confluence), and continuous build and deployment (Gitlab). To modularize implementation, the GitFlow approach was used. To separate continuous building and release versions, deployment was split up into a development, test, and production environment. Teams collaborated efficiently using Slack. For components of specific studies, the core development team was extended by internship students.

The app comprises a Web-based user interface and an object-oriented programming language (Java)–hypertext preprocessor (PHP) back end. The platform is set up on a cloud server. The front and back ends are coded in various languages. The front end is built with PHP with a combination of bootstrap and Laravel frameworks. The back end is built with a combination of PHP and Java to ensure cross-platform compatibility. The server runs on Windows, Apache, MySQL, and PHP platforms. Owing to the highly valuable content of the database system, the database is backed up daily and archived.

Technical Architecture and Data Security

The platform was built as a pluggable component-based middleware. As mentioned by Piwek et al [17], data security and patient privacy are essential to the adoption of digital smartphone research methods. A centralized data structure and shared research platform for multiple studies, as proposed in this study, eliminates the need to develop individual data security solutions for individual studies [18]. Given the use of highly sensitive data, the Health Insurance Portability and Accountability Act, especially the technical regulations, was used to achieve the highest possible level of security. We also followed the privacy by design principle of the general data protection regulation (GDPR). Strict security protocols are in place to ensure data safety, including several firewalls and secure sockets layer certificates embedded in the cloud-hosting infrastructure, virtual server, and database. OpenVPN is implemented to ensure a safe connection between the personal computer (PC) and the server. The database is encrypted by cipher block chaining in combination with advanced encryption standard (AES). User authentication for access to the platform is handled by a user login protected by Google's reCAPTCHA technique. All false login attempts are logged in the database, revealing unwanted access attempts and allowing us to create internet protocol exclusion rules. Communication between the database and the platform is encrypted by the AES-Rijndael algorithm. All data exchange with external databases is performed using the handshake principle, based on a standardized OAuth verification/authentication procedure.

Components

The concept of components was introduced to manage different studies on a single platform. Each study comprises at least 1 or multiple components (eg, data handling and questionnaires) that can be activated by a study leader.

Study and Patient Administration

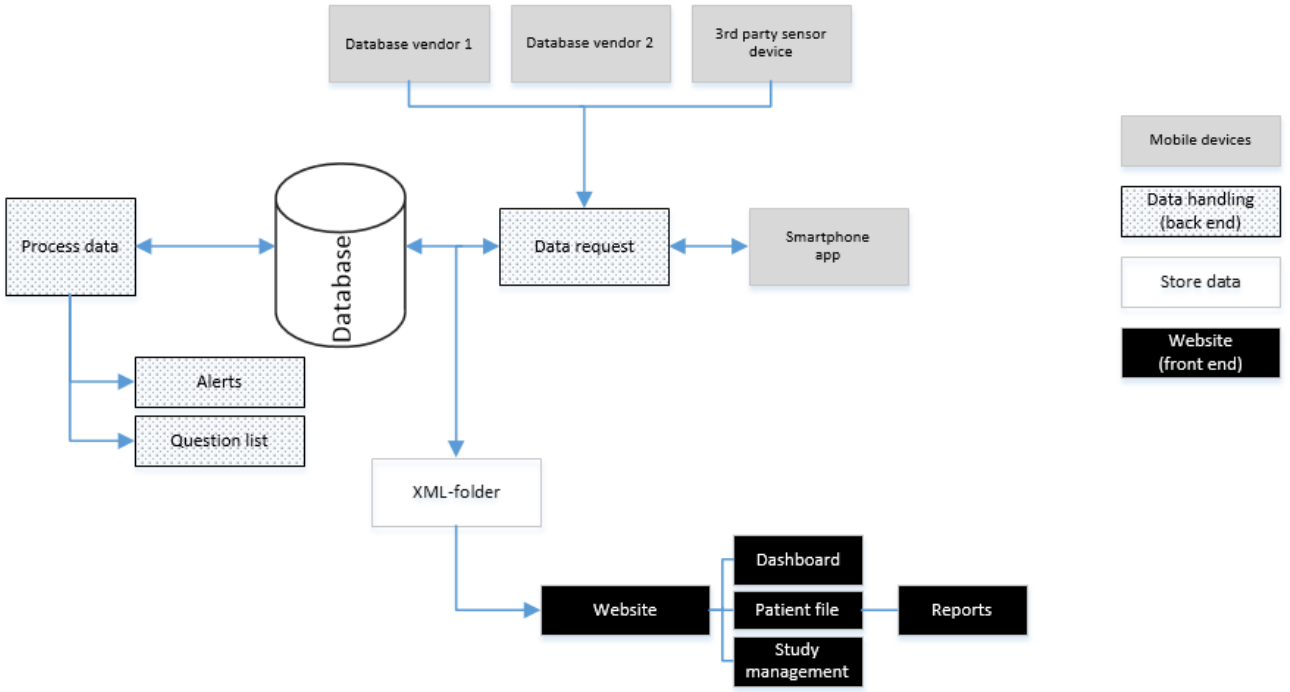
Every clinical study is divided into multiple levels, starting with the hospital acting as the lead partner of the clinical study (level 1; eg, Hospital East-Limburg, Genk, Belgium). The next level is the medical domain (level 2; eg, gynecology), and the third level is the study itself (level 3; eg, pregnancy remote monitoring study, Premom). This multilevel approach allows us to implement user rights and develop alert thresholds on each level of the study. Currently, 3 user profiles are defined: patient, study leader, and administrator (admin). After completing the registration process, 2 possible methods are provided. First, the admin can appoint a user as a study leader, and then a study leader can assign the user as a patient for inclusion in a specific study. This hierarchical model allows study leaders to independently create and follow-up patients.

Data Handling

The goal of DHARMA is to aggregate and visualize multiple vital parameters collected by using devices and medical apps from different vendors without the need to consult the vendor-specific platforms. The remote monitoring platform can receive information directly or collect data by connecting to other databases. If new values are uploaded by a user, DHARMA receives a notification and automatically launches the technical process needed to aggregate and store the data in its own database. Duplication of these data enables secure storage and accessibility for analysis and alert generation. Figure 1 is a graphic representation of the data flow in DHARMA. As the sensors record diverse types of data, ranging from discrete values to longitudinal values collected throughout the day, a metamodel was designed to handle and store data from each sensor. To handle large amounts of data, such as intraday (minute-by-minute) results from an activity tracker, data are compressed into a tailored XML file that is stored in a folder structure on the server.

To generate an efficient data collection workflow, an alert engine was developed, which can interpret and handle medical and technical alerts. Medical alerts are based on the collected data and detect a value outside the specified thresholds. The thresholds are specified for each study (level 3) and can dynamically support longitudinal changes and more complex interpretations. However, the configuration of patient-level alerts can be set to individual ranges based on clinical guidelines. On the basis of the clinical input, the alerts are categorized as normal, medium, or high priority. Technical alerts are defined by messages containing information about missed data transmissions. This triage system could help organize the clinical call center activities.

Figure 1. Overview of the data flow of the digital health research platform for mobile health platform.



Question Management

The *question management* component was implemented to receive additional, context-related information from patients via questionnaires. Context information could help researchers interpret (different) vital parameters [19]. Web- and app-based questionnaires can be developed and linked to DHARMA. For Web-based questionnaires, the URL for the questionnaire is sent to the patient's email address, and the patient can log in to the platform to answer the questionnaire. We developed the DHARMA smartphone app for app-based questionnaires. This app was built with the cross-platform Xamarin language to enable a solution compatible with both Android and iOS. The user interface was developed with Xamarin.Forms. A secured oauth2 app programming interface (API) connection between the platform and the app was developed to handle data flow. A local smartphone SQLite database is used to save and answer questionnaires offline. The app was integrated with firebase cloud messaging to notify patients when a questionnaire is delivered to the smartphone app. Questionnaires can be sent automatically via email or via the smartphone app by the Laravel task scheduler. This script comprises the questionnaire identifier (ID), study name, and date and time stamps. Questionnaires can be developed by the study leaders or admins and comprise individual questions. Each question has an identical ID and can be used in multiple questionnaires. Questions can be written as open questions, multiple choice questions, yes/no questions, or scale.

Patient File

A patient file component was created to arrange the information into individual patient records. Each record comprises 4 main tabs: medical information, statistics, questionnaires, and follow-up. The medical information tab allows the study leader and the clinician to view study-specific patient parameters or comorbidities. Study parameters include the study-specific information needed to interpret the vital parameters. The statistics tab displays a graphical overview of each vital parameter. The questions tab provides an overview of the questionnaires that were sent to and completed by the patient. The follow-up tab allows caregivers to send text messages among multiple disciplines. Each patient contact is logged in this tab.

Reports

A report component was created to provide a comprehensive digital overview of the patient's status. The overview can be printed or downloaded and emailed to the patient's doctor or caregiver. Lava Charts (Google Chart API) was used to visualize the patient's data in charts and graphs.

Results

The remote monitoring study platform was built between February 2015 and July 2018. The custom-made remote monitoring platform was deployed to monitor patients outside the hospital in several studies, including multiple sclerosis, low back pain, and osteoporosis studies in which the patients' activity data (number of steps and intensity) were tracked.

Overview of Platform Data

Table 1 provides an overview of the main studies in which DHARMA was tested, together with the numbers of patients and the vital parameters. Each vital parameter displays the number of individual measurements that were uploaded during the study.

Table 1. Remote studies included in digital health research platform for mobile health.

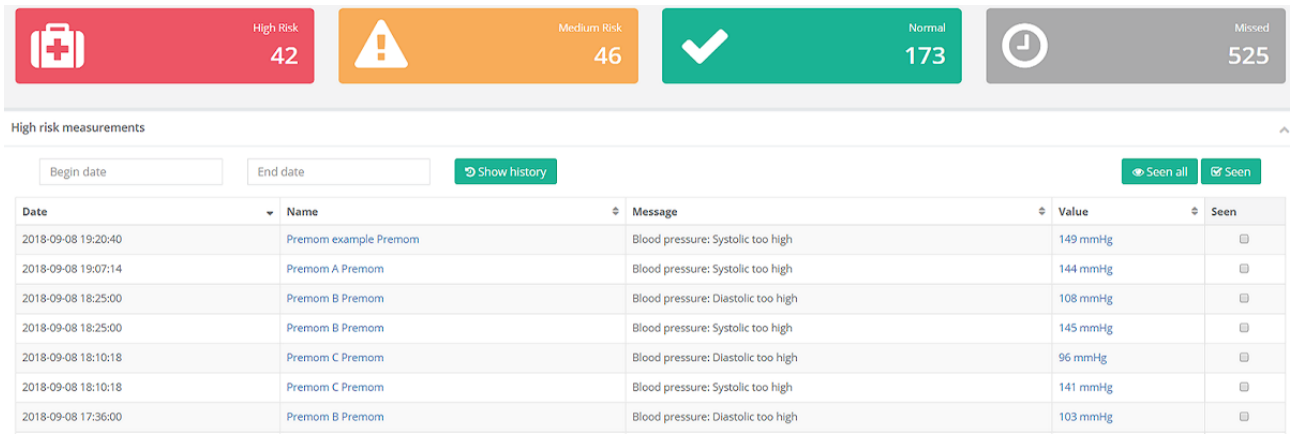
Study	Sample size, n	Blood pressure, n	Weight, n	Activity data, n
Premom	604	95,835	9430	35,520
Multiple sclerosis	36	0	14	1544
Low back pain	33	12	110	2191
Osteoporosis	13	25	857	1603

The main study in which DHARMA was tested was the Premom [20-22]. Briefly, this prospective cohort study enrolled pregnant women at high risk of developing pre-eclampsia. Patients were provided with a commercial activity tracker, wireless blood-pressure monitor, and a smart body scale analyzer from iHealth (iHealth Labs Inc, Mountain View, CA, USA) or Withings (Withings, Issy-les-Moulineux, France). The participating women were asked to measure their blood pressure twice daily, their weight once daily, and wear an activity tracker for 24 hours per day. All of the information recorded by the devices was sent wirelessly to the smartphone, which then transmitted the data to the Web-based platform for aggregation. The sensors collected up to 12 different signals as functions of time, enabling multiparametric longitudinal research. A midwife reviewed all incoming remote monitoring data via the dashboard. The alert engine discriminated between normal and alarm signals for the following: systolic blood pressure >140 mmHg, diastolic blood pressure >90 mmHg, or weight gain >1 kg per day. Alarm events were sent to the obstetrician to discuss possible interventions.

Dashboard and Visualization

Besides collecting and handling different sources of information, one of the main objectives of the platform is to provide the researcher/clinician with efficient visualization of all patient data. If the data triggered a specific alert, the dashboard prioritized the alerts based on the predefined thresholds and displayed them to the person responsible for reviewing the data. This enables the platform to triage patient alerts and facilitate patient handling and follow-up. Figure 2 shows an overview of the dashboard.

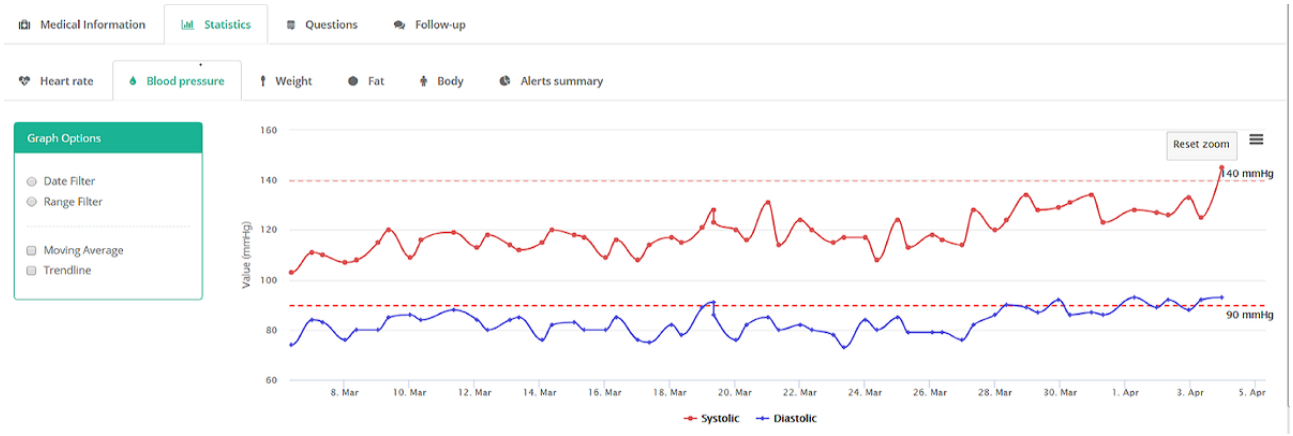
Figure 2. Screenshot of the alert representation upon login of the dashboard. Data was triaged based on predetermined thresholds into High Risk, Medium Risk or Normal.



Patient File

The patient records bundle the individual patient's information into a single file. The received parameters are individually plotted as functions of time to identify specific trends that could trigger an alert by crossing specified thresholds. For the patient shown in Figure 3, systolic blood pressure showed a trend toward crossing the predefined thresholds (140 mmHg for systolic blood pressure and 90 mmHg for diastolic blood pressure), triggering a high-risk alert. On the basis of these results, the patient was admitted to hospital where early symptoms of pre-eclampsia were identified, and appropriate treatment was started.

Figure 3. Overview graph of a patient's blood pressure displaying both the systolic and diastolic blood pressure, with the predefined alert thresholds indicated with dashed lines.



Data Structure and Data Handling: Activity Data

The challenges of working with different types of data include how to handle, analyze, and store the data appropriately and transparently. In the normal workflow, only the summarized values of larger datasets are used because granular detail is not required for daily patient management. However, if required, the data are available for Web-based data processing or can be exported for offline scientific research, such as the development of novel algorithms or data-processing techniques. Figure 4 shows the average step count over a period of 12 weeks. Granular, minute-by-minute data are also shown for a single day in Figure 5.

Figure 4. An example of longitudinal activity data for a period of about 2 months.

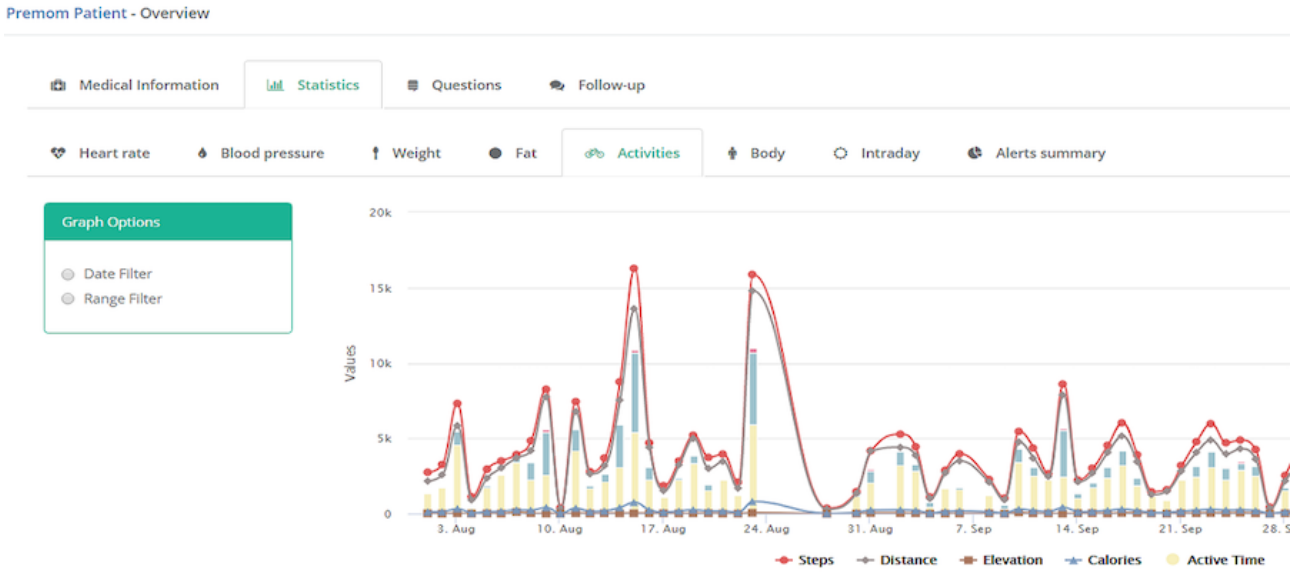
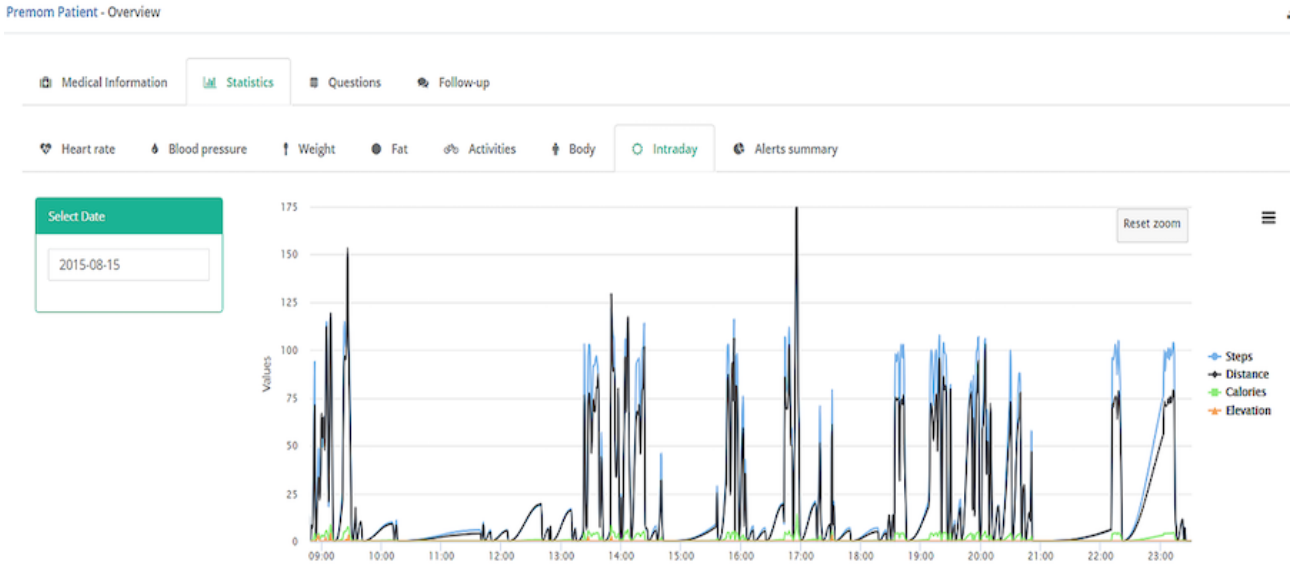


Figure 5. Granular example of the activity level per minute for a period of about 15 hours.



Discussion

Principal Findings

This paper outlines the development of a digital health research platform for remote monitoring. By combining advanced wearable sensors with smartphone technologies for remote monitoring, it is possible to monitor the health of patients in their home environment, an approach that may reduce the number of *health care visits*. Remote monitoring requires multiple hardware and infrastructure tools. Each vendor provides dedicated infrastructure and data review platforms specific to their own devices. Accordingly, data aggregation is impossible when collating data from medical devices and tools from different manufacturers, creating a barrier to clinical practice and academic research. This fragmentation is also very inconvenient for the user. Therefore, a *one size fits all* solution (ie, one platform for all devices) is highly desirable [23].

A health monitoring system can only provide its greatest usability if it can be fully integrated into the user's and the physician's daily workflow. The goal of our study is to integrate the data streams from multiple medical device vendors and allow health care practitioners to efficiently evaluate a patient's health status. This will improve efficiencies in cost and time. In addition, the platform was designed to enable rapid and cost-effective scalability.

Privacy is a fundamental right in the public health care domain, especially following the recently implemented GDPRs. Health care practitioners and patients are becoming increasingly aware of this important aspect. Confidential handling and storage of private patient data have also become a critical aspect of study design. Therefore, all personal data in our platform are deidentified and every unique identification number, characteristic, or name is removed. Moreover, all participants need to provide signed written informed consent.

Comparison With Previous Studies

The development of a centralized visualization platform has been described in earlier reports, for example, for monitoring arrhythmias [24], nonmotor symptoms of Parkinson disease [25], and pressure ulcers [26]. However, most of these studies monitored a specific disease, and thus the platforms have limited applicability to studies of other diseases. Zens et al [18] developed a modular smartphone app that could be used in different medical studies without the need for advanced programming skills. Another example of an open-source framework is ResearchKit (Apple Inc). However, the initial studies revealed that technical programming skills, such as Object C or Swift, are needed to develop a functional app [27-29]. Another limitation is that ResearchKit only supports iOS devices. Other platforms, including ResearchStack and ResearchDroid, have been developed for use in research projects. ResearchStack is a functional software development kit with a framework comparable to that of ResearchKit for developing research apps for Android devices [30]. ResearchDroid is an Android library developed to automate survey forms and the information-building process [31]. Appbakery integrates ResearchKit and ResearchDroid, enabling researchers to create apps without requiring programming skills [32]. More recently, Google's Open Data Kit has started allowing researchers to set up a study simply with a scalable app. Patient data can be exported to a comma-separated values file or viewed in the Google Cloud platform [33]. As examples of PC-based software, PsychoPy [34] and Labview (National Instruments) enable users to create individual software solutions with a graphical user interface in a process that does not require programming skills. The platform most similar to DHARMA was developed by Validic and can provide continuous access to personal health data obtained by over 350 in-home medical devices and wearables. Companies, such as Philips and IBM, also provide health platforms for remote monitoring. Although these commercially available platforms could have worked for the Premom research project, there were 3 main reasons why we chose to develop our own platform. First, because of the limited budget and the need for a vendor-independent research platform, we created our own solution that had the minimum number of components and required minimum development. Second, the platform needed to be flexible and customizable

for use with new study (invalidated) thresholds. Third, we used integrated components and functions that could differ from the development roadmap for commercial platforms.

Opportunities and Future Improvements

Although DHARMA provides exciting opportunities to improve remote monitoring services, it is not free from limitations. First, the data recorded by the medical devices are initially sent to the vendor's dedicated database. This means that the vendor (eg, iHealth or Withings) also owns the patient's data. An iHealth/Withings study profile without the patient's personal data was created in the Premom study to deidentify the patients who were included in the follow-up program. This approach could be improved by creating a third-party app that connects directly to the medical devices; however, not all vendors allow direct access to their medical device via an open API. This process would bypass data transfer to an external company. A second limitation is the applicability of remote monitoring studies among technophobic individuals and people with limited cognition or ability to express consent, such as neonates, elderly, and sedated patients in an intensive care unit [35]. Related to this limitation, some people may not comply with manual entry of daily information, especially in long-term monitoring settings [36]. Automated *invisible* wearables, such as smartwatches or smart clothing, could help address noncompliance issues. Finally, it is difficult to keep pace with the rapidly evolving smartphone and wearable sensor technologies. Our platform was initially developed using PHP version 5.6.18, whereas the latest version is 7.4 at the time of writing. An advantage of DHARMA is its flexible architecture that enables rapid integration of new smartphone and wearable sensor technologies as they become available.

Currently, participants in DHARMA remote monitoring studies need to provide written informed consent to be enrolled in each study, as previously reported by Eysenbach et al [37]. Zens et al [18] described an alternative approach that uses an eligibility module to check the inclusion criteria and an integrated electronic informed consent component to obtain consent via a customized app. In the future, a similar component could be integrated into the DHARMA mobile app.

Another step to improve the platform will involve embracing the definitions of standard information models and information technology communication standards, such as Health Level 7 fast healthcare interoperability resources, together with clinical terminologies, such as systematized nomenclature of medicine—clinical terms, to ensure interoperability with hospital electronic medical record (EMR) systems [38]. Our platform can be seamlessly integrated into a patient's daily life, but introducing it into a physician's standard workflow may require integration with existing EMR systems.

Conclusions

Smartphone health apps and medical devices collect large amounts of vendor-specific data. There are currently very few tools to collate and handle the data generated by multiple medical devices. We developed a component-based digital research platform to integrate the data in different formats from different sensors into a single integrated system. The platform performed well in a health care setting in real-time circumstances for the follow-up of pregnant women at risk of developing pre-eclampsia. The next stage in its development will involve integrating the platform with existing EMR systems to create a closed-loop information system.

Scientists or companies willing to contribute to this study are welcome to contact the authors.

Acknowledgments

This study is part of the Limburg Clinical Research Program UHasselt-ZOL-Jessa that is supported by the Limburg Sterk Merk Foundation, the province of Limburg, the Flemish government, Hasselt University, Ziekenhuis Oost-Limburg, and Jessa Hospital.

Conflicts of Interest

None declared.

References

1. Boulos MN, 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 5;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
2. Statista. 2018. Number of Smartphone Users Worldwide From 2014 to 2020 (in Billions) URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/
3. Research2Guidance. 2016. mHealth Economics 2016 – Current Status and Trends of the mHealth App Market URL: https://research2guidance.com/product/mhealth-app-developer-economics-2016/
4. Lee JH. Future of the smartphone for patients and healthcare providers. Healthc Inform Res 2016 Jan;22(1):1-2 [FREE Full text] [doi: 10.4258/hir.2016.22.1.1] [Medline: 26893944]
5. Baig MM, GholamHosseini H, Connolly MJ. Mobile healthcare applications: system design review, critical issues and challenges. Australas Phys Eng Sci Med 2015 Mar;38(1):23-38. [doi: 10.1007/s13246-014-0315-4] [Medline: 25476753]
6. Hamine SG, 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]

7. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2009 Jan;32(Suppl 1):S62-S67 [FREE Full text] [doi: 10.2337/dc09-S062] [Medline: 19118289]
8. Ramani GV, Uber PA, Mehra MR. Chronic heart failure: contemporary diagnosis and management. Mayo Clin Proc 2010 Feb;85(2):180-195 [FREE Full text] [doi: 10.4065/mcp.2009.0494] [Medline: 20118395]
9. Levy S. Atrial fibrillation, the arrhythmia of the elderly, causes and associated conditions. Anadolu Kardiyol Derg 2002 Mar;2(1):55-60 [FREE Full text] [Medline: 12101796]
10. Kruse CS, Mileski M, Moreno J. Mobile health solutions for the aging population: a systematic narrative analysis. J Telemed Telecare 2017 May;23(4):439-451. [doi: 10.1177/1357633X16649790] [Medline: 27255207]
11. Vashist SK, Schneider EM, Luong JH. Commercial smartphone-based devices and smart applications for personalized healthcare monitoring and management. Diagnostics (Basel) 2014 Aug 18;4(3):104-128 [FREE Full text] [doi: 10.3390/diagnostics4030104] [Medline: 26852680]
12. Brucal SG, Clamor GK, Pasiliao LA, Soriano JP, Varilla LP. Portable Electrocardiogram Device Using Android Smartphone. In: Proceedings of the 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 p. 104-128. [doi: 10.1109/EMBC.2016.7590751]
13. Zhang Y, Liu H, Su X, Jiang P, Wei D. Remote mobile health monitoring system based on smart phone and browser/server structure. J Healthc Eng 2015;6(4):717-737 [FREE Full text] [doi: 10.1260/2040-2295.6.4.717] [Medline: 27010652]
14. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? J Am Med Assoc 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
15. Yoo S, Kim B, Park H, Choi J, Chun J. Realization of real-time clinical data integration using advanced database technology. AMIA Annu Symp Proc 2003;2003:738-742 [FREE Full text] [Medline: 14728271]
16. Stead WW, Miller RA, Musen MA, Hersh WR. Integration and beyond: linking information from disparate sources and into workflow. J Am Med Inform Assoc 2000;7(2):135-145 [FREE Full text] [doi: 10.1136/jamia.2000.0070135] [Medline: 10730596]
17. Piwek L, Ellis DA. Can programming frameworks bring smartphones into the mainstream of psychological science? Front Psychol 2016;7:1252 [FREE Full text] [doi: 10.3389/fpsyg.2016.01252] [Medline: 27602010]
18. Zens M, Grotejohann B, Tassoni A, Duttenhoefer F, Südkamp NP, Niemeyer P. Development of a modular research platform to create medical observational studies for mobile devices. JMIR Res Protoc 2017 May 23;6(5):e99 [FREE Full text] [doi: 10.2196/resprot.7705] [Medline: 28536095]
19. Koohpayehzadeh J, Ahmadi MH, Dehnad A, Arabshahi SK, Bigdeli S, Yadollahi S. Validity and reliability of activities coaching context questionnaire. Med J Islam Repub Iran 2014;28:41 [FREE Full text] [Medline: 25405107]
20. Lanssens D, Vandenberk T, Smeets CJ, de Cannière H, Molenberghs G, van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. JMIR Mhealth Uhealth 2017 Mar 9;5(3):e25 [FREE Full text] [doi: 10.2196/mhealth.6552] [Medline: 28279948]
21. Lanssens D, Vonck S, Storms V, Thijs IM, Grieten L, Gyselaers W. The impact of a remote monitoring program on the prenatal follow-up of women with gestational hypertensive disorders. Eur J Obstet Gynecol Reprod Biol 2018 Apr;223:72-78 [FREE Full text] [doi: 10.1016/j.ejogrb.2018.02.015] [Medline: 29500948]
22. Lanssens D, Vandenberk T, Smeets CJ, de Cannière H, Vonck S, Claessens J, et al. Prenatal remote monitoring of women with gestational hypertensive diseases: cost analysis. J Med Internet Res 2018 Dec 26;20(3):e102 [FREE Full text] [doi: 10.2196/jmir.9552] [Medline: 29581094]
23. Weiß JP, Welzel T, Hartmann BJ, Hübner U, Teuteberg F. Towards designing a secure exchange platform for diabetes monitoring and therapy. Stud Health Technol Inform 2018;248:239-246. [doi: 10.3233/978-1-61499-858-7-239] [Medline: 29726443]
24. Raj S, Ray KC. A personalized arrhythmia monitoring platform. Sci Rep 2018 Jul 30;8(1):11395 [FREE Full text] [doi: 10.1038/s41598-018-29690-2] [Medline: 30061754]
25. Cancela J, Mascato SV, Gatsios D, Rigas G, Marcante A, Gentile G, et al. Monitoring of Motor and Non-Motor Symptoms of Parkinson's Disease Through a mHealth Platform. In: Proceedings of the 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 p. 663-666. [doi: 10.1109/EMBC.2016.7590789]
26. Rodrigues JJ, Pedro LM, Vardasca T, de la Torre-Díez I, Martins HM. Mobile health platform for pressure ulcer monitoring with electronic health record integration. Health Informatics J 2013 Dec;19(4):300-311. [doi: 10.1177/1460458212474909] [Medline: 24255053]
27. Webster DE, Suver C, Doerr M, Mounts E, Domenico L, Petrie T, et al. The Mole Mapper study, mobile phone skin imaging and melanoma risk data collected using ResearchKit. Sci Data 2017 Dec 14;4:170005 [FREE Full text] [doi: 10.1038/sdata.2017.5] [Medline: 28195576]
28. Zens M, Woias P, Suedkamp NP, Niemeyer P. 'Back on Track': a mobile app observational study using Apple's ResearchKit framework. JMIR Mhealth Uhealth 2017 Feb 28;5(2):e23 [FREE Full text] [doi: 10.2196/mhealth.6259] [Medline: 28246069]
29. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The asthma mobile health study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]

30. Estrin DC. GitHub. 2017. An SDK for Building Research Study Apps on Android URL: https://github.com/ResearchStack/ResearchStack
31. Suhail M. Applied Informatics. 2016. ResearchDroid: An Android Forms and Consent Library URL: http://blog.appliedinformaticsinc.com/researchdroid-an-android-forms-and-consent-library/
32. TrialX: Appbakery - Mobilize Research. 2015. URL: http://trialx.com/appbakery/
33. Sharif B, Lundin RM, Morgan P, Hall JE, Dhadda A, Mann C, Welsh Digital Data Collection Platform Collaborators (see Appendix A). Developing a digital data collection platform to measure the prevalence of sepsis in Wales. J Am Med Inform Assoc 2016 Dec;23(6):1185-1189. [doi: 10.1093/jamia/ocv208] [Medline: 27094989]
34. Peirce JW. PsychoPy--psychophysics software in python. J Neurosci Methods 2007 May 15;162(1-2):8-13 [FREE Full text] [doi: 10.1016/j.jneumeth.2006.11.017] [Medline: 17254636]
35. Yang G, Jiang M, Ouyang W, Ji G, Xie H, Rahmani AM, et al. IoT-based remote pain monitoring system: from device to cloud platform. IEEE J Biomed Health Inform 2018 Dec;22(6):1711-1719. [doi: 10.1109/JBHI.2017.2776351] [Medline: 29990259]
36. Jonassaint CR, Shah N, Jonassaint J, de Castro L. Usability and feasibility of an mhealth intervention for monitoring and managing pain symptoms in sickle cell disease: the sickle cell disease mobile application to record symptoms via technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]
37. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
38. Benson T. Using SNOMED and HL7 together. In: Principles of Health Interoperability HL7 and SNOMED. London: Springer; 2019:267-280.

Abbreviations

AES: advanced encryption standard
API: app programming interface
DHARMA: digital health research platform for mobile health
ECG: electrocardiogram
EMR: electronic medical record
GDPR: general data protection regulation
PC: personal computer

Edited by CL Parra-Calderón; submitted 23.10.18; peer-reviewed by M Shanmugham, JH Lee; comments to author 17.12.18; revised version received 22.12.18; accepted 15.06.19; published 29.10.19.

Please cite as:
Vandenberk T, Storms V, Lanssens D, De Cannière H, Smeets CJP, Thijs IM, Batool T, Vanrompay Y, Vandervoort PM, Grieten L
A Vendor-Independent Mobile Health Monitoring Platform for Digital Health Studies: Development and Usability Study
JMIR Mhealth Uhealth 2019;7(10):e12586
URL: https://mhealth.jmir.org/2019/10/e12586
doi:10.2196/12586
PMID:31663862

Review

Investigating the Use of Mobile Health Interventions in Vulnerable Populations for Cardiovascular Disease Management: Scoping Review

Sahr Wali[1,2], MSc; Neesha Hussain-Shamsy[1], MSc; Heather Ross[3,4,5], MHSc, MD, FRCPC; Joseph Cafazzo[1,2,6], PEng, PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[3]Ted Rogers Centre for Heart Research, University Health Network, Toronto, ON, Canada

[4]Department of Medicine, University of Toronto, Toronto, ON, Canada

[5]Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada

[6]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Sahr Wali, MSc
Centre for Global eHealth Innovation
Techna Institute
University Health Network
TGH/RFE Bldg, 4th Fl
190 Elizabeth St
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 3200
Email: sahr.wali@mail.utoronto.ca

Abstract

Background: Cardiovascular disease (CVD) has grown to become one of the leading causes of mortality worldwide. The advancements of CVD-related treatments have led to a decline in CVD prevalence among individuals in high-income countries (HICs). However, these improvements do not reflect the state of individuals in low- and middle-income countries (LMICs) and vulnerable subgroup populations in HICs, such as the Indigenous. To help minimize the health disparities in these populations, technology-based interventions have been offered as a potential solution, but there is concern regarding if they will be effective, or even needed, as these tools have been designed for use in HICs.

Objective: The objective of this study was to explore how mobile health (mHealth) interventions currently assist individuals in Indigenous communities and LMICs with CVD management.

Methods: A scoping review guided by the methods outlined by Arksey and O'Malley was conducted. A comprehensive search was completed by 2 reviewers in 5 electronic databases using keywords related to mobile health, cardiovascular disease, self-care, Indigenous communities, and LMICs. Studies were screened over 2 rounds and critically reviewed using a descriptive-analytical narrative method. Descriptive data were categorized into thematic groups reflecting the major findings related to the study objective.

Results: We identified a total of 11 original articles and 11 review papers that met the criteria for this scoping review. The majority of the studies included a telemonitoring- and text messaging (short message service, SMS)–related feature associated with the intervention. The use of SMS was the most common approach to effectively promote disease management among individuals in both LMICs and Indigenous communities. However, customizing for cultural considerations within the design of the intervention was highlighted as a pivotal component to encourage CVD management. Specifically, individuals emphasized that the inclusion of collaborative partnerships with community members would strengthen the effectiveness of the intervention by ensuring it was designed with the appropriate context.

Conclusions: Technology-based interventions used within Indigenous communities and LMICs have shown their potential to assist individuals with managing their condition. Although the literature available regarding this topic is limited, this review outlines key components to promote the effective use of these tools in the context of these vulnerable populations.

(JMIR Mhealth Uhealth 2019;7(10):e14275) doi:10.2196/14275

KEYWORDS

mobile health; health services; indigenous; low- and middle-income countries; cardiovascular disease; self-care

Introduction

The World Health Organization has indicated that cardiovascular disease (CVD) is the leading cause of mortality, accounting for nearly 31% of all deaths worldwide [1-3]. Despite the increase in prevalence of the disease, CVD is largely treatable or controllable if patients are able to properly manage and monitor their symptoms [2,4]. Previous studies have found that the promotion of CVD self-care led to improved cardiovascular outcomes [2-4]. In Canada, the dissemination of CVD self-care strategies and clinical practice guidelines led the rate of CVD to decline among most age groups [4,5]. However, these improvements did not reflect the state of more vulnerable and at-risk populations within the country, such as the Indigenous [4,5].

Indigenous populations across various high-income countries (HICs), such as Canada, the United States, and Australia, have reported that they have experienced a rise in CVD prevalence and associated mortality, relative to the rest of the population [3-6]. The discrepancy in CVD improvement across HICs is largely linked to poor disease control and disparities in social determinants of health [2,4,7]. This includes factors such as housing, education, access to health services, and income [4].

Many of the risk factors affecting Indigenous populations are shared by individuals from low- and middle-income countries (LMICs) [4,8]. The effect of these conditions is supplemented by the lack of preventative strategies and health care resources available for both populations [2,3]. Indigenous populations do not receive or have access to the same level of care for the treatment or management of CVD as non-Indigenous populations [9]. Consequently, they are 3 times more likely to die from a CVD-related event [4,9]. Similarly, there are currently 17.5 million CVD-related deaths occurring annually, and more than 75% of the CVD-related deaths are accounted by LMICs because of the lack of clinics and assistive tools available for appropriate diagnosis and care [8]. Both these populations may have differences in their history, culture, and structure of their health care system, but the underlying shared causes leading to poor patient outcomes (ie, geographic and socioeconomic factors) create a comparable link between the 2 populations. Thus, as both Indigenous communities and individuals with CVD living in LMICs face challenges involving accessibility to appropriate resources, more innovative interventions promoting CVD management within their environmental setting need to be introduced [4,10].

With the widespread penetration of mobile phones, mobile health (mHealth) technology offers a promising platform to support CVD management in populations with less health resources [8,11]. It is estimated that more than 60% of individuals in LMICs own a mobile phone and up to 85% have access to a device [12]. mHealth apps could serve as a low-cost tool to simplify and promote CVD management [13-15]. Previous studies have shown that mHealth tools and other technology-related interventions have had a positive impact on improving patient outcomes [13-16]. This has been supplemented by its ability to simplify or automate self-care steps as well as provide additional support and guidance from a health care professional [14-16]. However, as the majority of mHealth interventions have been designed for use in HICs, there is concern regarding whether these tools will be effective in low health resource settings [15]. To address this gap in knowledge, this review aims to study how various mHealth interventions (ie, apps, text message, and telemonitoring) assist Indigenous communities and individuals in LMICs with CVD management (ie, self-care and remote management).

Methods

Review Framework

This review was guided by Arksey and O'Malley's 5-stage scoping review framework: (1) identifying the research question, (2) identifying relevant studies, (3) selection of studies, (4) charting the data, and (5) summarizing and reporting the results [17]. Their review approach synthesizes and maps key concepts from the literature available to give a better understanding of the impact of innovative self-care tools in more vulnerable populations.

Research Question

The focus of this review was to explore how mHealth interventions for CVD management assisted individuals from Indigenous populations and LMICs. mHealth tools are often evaluated in HICs; thus, this led to the following guiding question: *What is known in the literature about the use of mHealth interventions on CVD management in Indigenous communities and LMICs?*

Search Strategy

A preliminary scan of literature was conducted on 2 academic databases (MEDLINE and EMBASE) with the following search terms: mobile health, cardiovascular disease, self-care, Indigenous, and LMIC. Keywords and related subject headings were refined according to text contained in the title and abstract of the initial literature search (Textbox 1). On the basis of the keywords identified, 2 reviewers (SW and NHS) independently conducted a comprehensive search in 5 electronic databases: MEDLINE, EMBASE, Web of Science, Cumulative Index to Nursing and Allied Health Literature, and Scopus. Reference lists were also reviewed to extract additional studies that were not found in the initial search. This review did not restrict studies according to the year of publication, but studies were considered ineligible if they were not published in English.

Textbox 1. Scoping review search strategy.

Scoping review keywords:

1. Mobile health

- Mobile health OR mHealth OR digital health OR health application* OR health app* OR health technolog* OR eHealth OR mobile phone* OR phone-based OR SMS OR short message service* OR text message* OR telehealth OR telephone monitor* OR telemedicine

AND

2. Cardiovascular disease

- Cardiovascular disease* OR CVD OR heart failure OR HF OR stroke* OR heart attack* OR chronic disease* OR hypertension OR HT

AND

3. Self-care

- Self-car* OR disease manag* OR self manag* OR remote manag* OR remote car* OR manag* OR treatment

AND

4a. Indigenous

- Indigenous OR Aboriginal* or Metis OR Inuit* OR First Nation* or Native American*

OR

4b. LMIC

- LMIC OR low income OR middle income OR developing countr* OR third world countr* OR vulnerable population*)

Study Selection

The retrieved literature was screened over 2 rounds for study selection. In the first round, titles and abstracts were reviewed according to the inclusion and exclusion criteria listed in Textboxes 2 and 3, respectively. In the second round, abstracts and the full text were screened to determine if they met the outlined criteria.

Textbox 2. Inclusion criteria for study selection.

Inclusion criteria:

- Primary intervention involves mobile health– or technology-related tool or aid for self-care
- Distinction of an Indigenous or low- and middle-income country population—study could include nonvulnerable population in addition to Indigenous population or individuals in low- and middle-income country
- Study population has at least one cardiovascular disease–related condition (ie, heart failure, stroke, or heart attack)

Textbox 3. Exclusion criteria for study selection.

Exclusion criteria:

- Intervention does not include cardiovascular disease management, monitoring, or promotion as a key component in the study
- Primary study population is high-income country
- Grey literature, review papers, or study protocols

Charting and Extracting Data

Articles meeting the inclusion criteria were critically reviewed using Arksey and O'Malley's descriptive-analytical narrative method [16,17]. Data extracted included the year of publication, study location, intervention type (phone, tablet, telemedicine, etc), study population, aim of study, methodology, outcome measures, and main findings.

Summarizing and Reporting Results

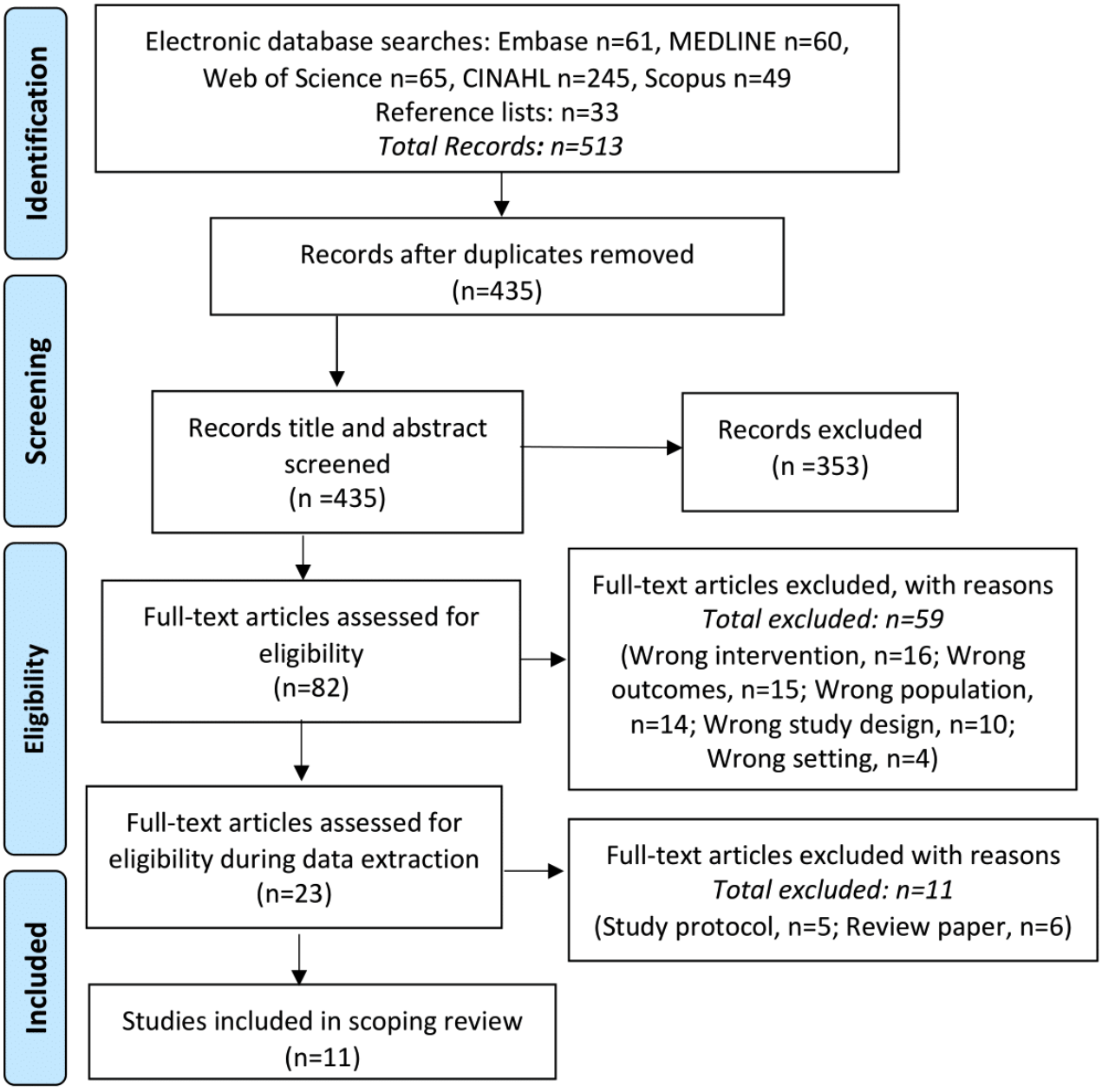
A numerical analysis of the extent and nature of the studies was conducted using tables and chart mappings. The descriptive data were analyzed using conventional content analysis. In accordance with the user-centered design framework, 2 reviewers (SW and NHS) examined the descriptive data and identified codes relative to the findings [16]. These codes were then organized into thematic groups to summarize the literature according to their main findings and present a narrative relating to the research question.

Results

Common Themes

A total of 513 articles were identified from the 5 databases and reference lists searched. Furthermore, 78 duplicate articles were removed, and the remaining 435 articles were screened according to the inclusion and exclusion criteria listed in Textboxes 2 and 3, respectively. After a review of the title and abstracts, 82 articles were included for the full-text review. Moreover, 59 articles were excluded during the full-text screening because of a series of implications with the inclusion and exclusion criteria outlined in Figure 1. A total of 11 articles were included in the scoping review (Figure 1). However, owing to the limited amount of studies found, the review papers and study protocols meeting the inclusion criteria were reviewed separately for common themes.

Figure 1. Systematic scoping review search strategy—Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart. CINAHL: Cumulative Index to Nursing and Allied Health Literature.



Findings From Original Studies

From the 11 articles, the most common study locations were Latin America (number of articles [n]=3), New Zealand (n=2), and South Africa (n=2). However, the majority of the study populations were LMICs (n=8; Multimedia Appendix 1). The interventions used within various studies were differentiated by the combination of their use of (1) short message service (SMS) text messaging and telemonitoring (phone calls; n=3), (2) telemonitoring only (n=4), and (3) SMS text messages only [3]. An article did not have an intervention outlined as it was a qualitative study focused on identifying the barriers and opportunities for technology in Indigenous populations [18].

The aim of 10 of these studies was to evaluate the outlined interventions' ability to improve disease management. This

included the impact of a reminder system (n=6), health education (n=2), and telemonitoring (n=6). A total of 8 of the studies commonly used a behavior change or the satisfaction of using the intervention as their primary outcome measure (Multimedia Appendix 1).

The main findings from these studies indicated that the various mHealth interventions promoted a positive behavior change for CVD management. This included improved medication adherence (n=4), monitoring and control of systolic blood pressure (BP; n=4), and overall self-care (n=5). Of the articles, 6 included an SMS component to the intervention, and in 5 of those studies, patients explicitly stated that they preferred the SMS text messages as a reminder system. In 2 studies, patients using the self-care program reported improvements in their satisfaction with their care and overall quality of life compared with the control group.

Findings From Review Papers or Study Protocols

Of the selected articles, 6 were review papers (4 systematic reviews and 1 literature review and 1 rapid review), 4 were randomized controlled trial (RCT) protocols, and 1 was a co-design paper (Multimedia Appendix 2, [19-21]). The majority of these articles' study locations were in Australia (n=5), and of the 11 articles, there was an even distribution between the study populations (5 LMIC, 5 Indigenous, and 1 LMIC and Indigenous).

The interventions included within the articles varied by the following components: (1) mHealth app and SMS text message (n=4), (2) mHealth app and educational program (n=1), (3) mHealth app and telemonitoring (n=3), and (4) mHealth app only (n=3). The aim of these articles was to provide evidence on the effect mobile phone–based solutions had on CVD management within either population (n=6). Out of 4 of the review articles, 3 found that mHealth tools were a sufficient method to promote CVD management, specifically with the use of SMS text messages. In 3 articles, community collaboration and cultural sensitivity were also highlighted as key components for intervention success.

Findings From Thematic Analysis

In total, 2 dominant themes related to the design of mHealth intervention and its resultant effect were observed: (1) establishing a reminder system and (2) customizing for cultural considerations.

Establishing a Reminder System

Medication management was found to be the most common form of CVD-related self-care in LMICs and Indigenous communities [22-32]. In 2016, Gu et al [18] found that the issues contributing to poor medication adherence in the Australian Indigenous communities included poor patient knowledge of medication effects, cost-value of medication use, and general forgetfulness. Multiple studies indicated that mHealth solutions promoted medication adherence by acting as a reminder system [22-32].

Specifically, the combination of SMS text messages and telemonitoring was shown to be effective in improving BP control and overall treatment adherence [22,24]. In a study by Sarfo et al [33], they introduced a phone-based intervention that included Bluetooth telemonitoring of BP, SMS text messages, and nurse guidance for HF self-care. After 9 months of the intervention's implementation, patients in the intervention group reported higher medication adherence scores and better systolic BP control (73.3% of patients with systolic B*P*<140 mmHg) compared with patients in the control group (43.3% of patients with systolic B*P*<140 mmHg) [33].

Telemonitoring alone also had a significant effect on improving disease management [26-28]. In 2012, Piette et al conducted an RCT to test the effectiveness of a self-management system that included telemonitoring, behavior change calls, and home BP monitoring [28]. They found that patients receiving the intervention had a reduction of 4.2 mmHg in their BP compared with the control group [28]. Patients had also reported that the intervention increased their medication adherence and satisfaction with their care. In 2016, Piette et al tested the same platform in Bolivia with the addition of a CarePartner feature, where a family or friend would receive a summary of the patient's status and guidance on how to support their self-management [26]. Patients receiving the intervention had significant improvements in their self-care reports, especially among Indigenous and low-literacy patients [26].

The use of SMS text messages alone was noted as the most common and effective method to increase medication adherence in LMICs compared with smartphone apps [34,35]. In a study by Kamal et al [24], they evaluated the effect of SMS medication reminders on recent stroke survivors in Pakistan. Results indicated that patients receiving the intervention had a 4.09 times lower risk of being low adherent [24]. In a study by Hacking et al [23], they found similar results, as SMS text messages improved self-reported CVD management. In addition, their study found that patients preferred the SMS text messages as a reminder system more than education.

Customizing for Cultural Considerations

Among the target populations, the inclusion of unique cultural considerations was identified as a key component for the promotion of CVD management [35,36]. In 2016, Gu et al [18] conducted a participatory action focus group, and they identified that technology-based solutions in Indigenous communities would need to be both culturally and literacy sensitive. Banbury et al [37] explained that Indigenous populations already had a *preset fear of nontraditional tools*; therefore, new technologies would need to be developed in partnership with the community to ensure they reflected the populations' needs.

Bradford et al [9] collaborated with the Mayor and Council of a remote Aboriginal Community to customize their existing mobile phone cardiac rehabilitation health program for Indigenous Australians. Following this study, the rehabilitation program was modified by simplifying the text and user interface, adding Indigenous artwork, and framing educational material in a culturally appropriate context.

Discussion

Principal Findings

The purpose of this scoping review was to identify how mHealth interventions were utilized in Indigenous populations and individuals in LMICs for CVD management. The amount of available literature exploring this topic was very limited as the target population for most studies was based in an HIC. However, based on the articles found, the use of mHealth solutions showed the potential to improve CVD management through the promotion of medication adherence and health education.

Throughout this review, more than half of the patients in the studies identified were classified as nonadherent [26]. The majority of Indigenous people and LMIC populations reside in remote areas that are limited in the health resources available [4,37]. The addition of telemonitoring and/or SMS text messaging allows patients unable to attend obtain medical assistance to actively manage their treatment at home [26-28].

The use of SMS text messages was more common than telemonitoring or other interactive methods within the target populations [22,23]. Regardless of its prevalence, phone calls were identified as a better alternative to promote disease management because of its ability to cater to patients with low health literacy and serve as more than just a reminder system by providing additional education [25-28]. With this, there is a concern that the use of SMS may not be as effective in promoting self-care as anticipated. However, in both SMS- and phone call–based interventions, the studies may have reported that medication adherence and patient self-care improved, but their effect size was only quantified in 1 study [24]. To determine the impact of each intervention on CVD management and other key outcome measures, further research is required.

The beneficial role of a caregiver, in addition to the use of an mHealth intervention, was also highlighted in this review. In a series of 3 studies conducted by Piette et al, they had tested the use of their self-management system (ie, telemonitoring, phone calls, and home BP management) and found that the addition of a CarePartner significantly improved patient self-care compared with the use of the mHealth intervention alone or standard care [26-28]. These findings support the idea that by collaborating with carers and partners within a self-care program, patient disease management can be improved. The literature supporting this topic for these specific populations is limited, but it provides an additional avenue for further study.

A key strength the studies highlighted involved their identification of cultural factors that needed to be considered within their design. Specifically, individuals in both Indigenous populations and LMICs commonly have low literacy levels. Their lack of knowledge or understanding of their medication results in compromising their ability to follow their treatment regimen. A few of the studies found were in the process of identifying factors that would need to be accommodated for within their intervention, and 1 study had already begun modifying their mHealth platform [9,18,36]. The importance of cultural factors was found to be more evident in Indigenous populations compared with LMICs. To further understand if there is a comparable link between both populations and their use of mHealth interventions for self-care, future research should look to explore what key determinants are responsible for promoting self-care.

Limitations

This scoping review was limited as most of the articles' results were composed of patients' self-reported data that might have been inaccurate. We also included review papers and protocol papers within this review to expand the breadth of studies analyzed, but this might have impacted the validity of our overall findings. In addition, only 3 out of the 11 original articles included an Indigenous study population. Moreover, 2 of those studies including the Indigenous population were not able to evaluate the impact of the intervention on patient outcomes, as one was a qualitative participatory action focus group and the other was focused solely on intervention feasibility.

Conclusions

The inequities present among Indigenous communities and LMICs contribute to the growing prevalence of CVD within their populations. mHealth interventions for disease management have primarily been implemented in HICs; however, simpler versions of these tools have been shown to improve CVD self-care in these more vulnerable populations. Cultural compatibility is essential for the success of these interventions in low health resource settings. Nevertheless, although the literature supporting mHealth in these populations is limited, their results indicate that they have a promising future.

Acknowledgments

The authors would like to thank Dr Whitney Berta for guiding the search strategy for this review. The authors would also like to recognize the previous study completed by Mr Jason Hearn that led to the development of this project.

Authors' Contributions

SW, HR, and JC contributed to the design of the review. SW and NHS reviewed, screened, and analyzed the extracted articles. SW drafted the paper for all the authors to comment and revise. All authors approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Charting of scoping review—original studies.
[PDF File (Adobe PDF File)153 KB - mhealth_v7i10e14275_app1.pdf]

Multimedia Appendix 2
Charting of scoping review—review and design methodology papers.
[PDF File (Adobe PDF File)122 KB - mhealth_v7i10e14275_app2.pdf]

References

1. World Health Organization. 2014. Global Status Report On Noncommunicable Diseases 2014 URL: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf;jsessionid=BB1EAD06225052FDC59ACEFA0B98E3F1?sequence=1 [accessed 2018-10-15]
2. Atzema CL, Khan S, Lu H, Allard YE, Russell SJ, Gravelle MR, et al. Cardiovascular disease rates, outcomes, and quality of care in Ontario Métis: a population-based cohort study. PLoS One 2015;10(3):e0121779 [FREE Full text] [doi: 10.1371/journal.pone.0121779] [Medline: 25793978]
3. Kandasamy S, Anand SS. Cardiovascular disease among women from vulnerable populations: a review. Can J Cardiol 2018 Apr;34(4):450-457. [doi: 10.1016/j.cjca.2018.01.017] [Medline: 29571426]
4. Yeates K, Campbell N, Maar MA, Perkins N, Liu P, Sleeth J, et al. The effectiveness of text messaging for detection and management of hypertension in indigenous people in Canada: protocol for a randomized controlled trial. JMIR Res Protoc 2017 Dec 19;6(12):e244 [FREE Full text] [doi: 10.2196/resprot.7139] [Medline: 29258978]
5. Prince SA, McDonnell LA, Turek MA, Visintini S, Nahwegahbow A, Kandasamy S, et al. The state of affairs for cardiovascular health research in indigenous women in Canada: a scoping review. Can J Cardiol 2018 Apr;34(4):437-449 [FREE Full text] [doi: 10.1016/j.cjca.2017.11.019] [Medline: 29439893]
6. Lucero AA, Lambrick DM, Faulkner JA, Fryer S, Tarrant MA, Poudevigne M, et al. Modifiable cardiovascular disease risk factors among indigenous populations. Adv Prev Med 2014;2014:547018 [FREE Full text] [doi: 10.1155/2014/547018] [Medline: 24649368]
7. Maar M, Yeates K, Barron M, Hua D, Liu P, Lum-Kwong MM, et al. I-RREACH: an engagement and assessment tool for improving implementation readiness of researchers, organizations and communities in complex interventions. Implement Sci 2015 May 4;10:64 [FREE Full text] [doi: 10.1186/s13012-015-0257-6] [Medline: 25935849]
8. Balakumar P, Maung- UK, Jagadeesh G. Prevalence and prevention of cardiovascular disease and diabetes mellitus. Pharmacol Res 2016 Nov;113(Pt A):600-609. [doi: 10.1016/j.phrs.2016.09.040] [Medline: 27697647]
9. Bradford DK, Hansen D, Karunanithi M. Making an appropriate care program for indigenous cardiac disease: customization of an existing cardiac rehabilitation program. Stud Health Technol Inform 2015;216:343-347. [Medline: 26262068]
10. World Health Organization. Global Health Observatory Data Repository URL: http://apps.who.int/gho/data/node.main.A865CARDIOVASCULAR?lang=en [accessed 2018-10-20]
11. Coorey GM, Neubeck L, Usherwood T, Peiris D, Parker S, Lau AY, et al. Implementation of a consumer-focused ehealth intervention for people with moderate-to-high cardiovascular disease risk: protocol for a mixed-methods process evaluation. BMJ Open 2017 Jan 11;7(1):e014353 [FREE Full text] [doi: 10.1136/bmjopen-2016-014353] [Medline: 28077414]
12. Bastawrous A, Armstrong MJ. Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. J R Soc Med 2013 Apr;106(4):130-142 [FREE Full text] [doi: 10.1177/0141076812472620] [Medline: 23564897]
13. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
14. 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]
15. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
16. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Gu Y, Kennely J, Warren J, Ahn AB, Harwood M, Neuwelt P. Identifying ehealth opportunities to support medication adherence - findings of a focus group study. Stud Health Technol Inform 2016;223:150-157. [doi: 10.3233/978-1-61499-645-3-150] [Medline: 27139398]

19. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018 Feb 7;18(1):25 [FREE Full text] [doi: 10.1186/s12872-018-0764-x] [Medline: 29415680]
20. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF, IMPACT Study Group. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192 [FREE Full text] [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]
21. Abrahams-Gessel S, Beratarrechea A, Irazola V, Gutierrez L, Moyano D, Fernandez A, et al. Using mhealth tools to improve access, coverage and treatment of uninsured people with high cardiovascular disease risk in Argentina: a study protocol for a pragmatic cluster randomised trial. BMJ Innov 2018 Jun 2;4(3):135-141. [doi: 10.1136/bmjinnov-2017-000255]
22. Bobrow K, Farmer AJ, 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 9;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
23. Hacking D, Haricharan HJ, Brittain K, Lau YK, Cassidy T, Heap M. Hypertension health promotion via text messaging at a community health center in South Africa: a mixed methods studyHypertension health promotion via text messaging at a community health center in south africa: a mixed methods study. JMIR Mhealth Uhealth 2016 Mar 10;4(1):e22 [FREE Full text] [doi: 10.2196/mhealth.4569] [Medline: 26964505]
24. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored short messaging service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
25. Nundy S, Razi RR, Dick JJ, Smith B, Mayo A, O'Connor A, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. J Med Internet Res 2013 Mar 11;15(3):e53 [FREE Full text] [doi: 10.2196/jmir.2317] [Medline: 23478028]
26. Piette JD, Marinec N, Janda K, Morgan E, Schantz K, Yujra AC, et al. Structured caregiver feedback enhances engagement and impact of mobile health support: a randomized trial in a lower-middle-income country. Telemed J E Health 2016 Apr;22(4):261-268 [FREE Full text] [doi: 10.1089/tmj.2015.0099] [Medline: 26352854]
27. Piette JD, Valverde H, Marinec N, Jantz R, Kamis K, de la Vega CL, et al. Establishing an independent mobile health program for chronic disease self-management support in Bolivia. Front Public Health 2014;2:95 [FREE Full text] [doi: 10.3389/fpubh.2014.00095] [Medline: 25165687]
28. Piette J, Lun K, Moura L, Fraser H, Mechael P, Powell J, et al. Impacts of e-health on the outcomes of care in low- and middle-income countries: where do we go from here? Bull World Health Organ 2012 May 1;90(5):365-372 [FREE Full text] [doi: 10.2471/BLT.11.099069] [Medline: 22589570]
29. Maslakpak MH, Safaie M. Effect of reminder cards on adherence to treatment in patients with uncontrolled hypertension. Iran J Nurs Res 2016 Dec 10;11(5):26-31. [doi: 10.21859/ijnr-11055]
30. Riley WT, Keberlein P, Sorenson G, Mohler S, Tye B, Ramirez AS, et al. Program evaluation of remote heart failure monitoring: healthcare utilization analysis in a rural regional medical center. Telemed J E Health 2015 Mar;21(3):157-162 [FREE Full text] [doi: 10.1089/tmj.2014.0093] [Medline: 25025239]
31. Venter A, Burns R, Hefford M, Ehrenberg N. Results of a telehealth-enabled chronic care management service to support people with long-term conditions at home. J Telemed Telecare 2012 Apr;18(3):172-175. [doi: 10.1258/jtt.2012.SFT112] [Medline: 22362838]
32. Alghamdi M, Gashgari H, Househ M. A systematic review of mobile health technology use in developing countries. Stud Health Technol Inform 2015;213:223-226. [Medline: 26152999]
33. Sarfo FS, Treiber F, Gebregziabher M, Adamu S, Nichols M, Singh A, PINGS Team. Phone-based intervention for blood pressure control among Ghanaian stroke survivors: a pilot randomized controlled trial. Int J Stroke 2018 Nov 22:- (epub ahead of print). [doi: 10.1177/1747493018816423] [Medline: 30465630]
34. Abaza H, Marschollek M. mHealth application areas and technology combinations*. A comparison of literature from high and low/middle income countries. Methods Inf Med 2017 Aug 8;56(7):e105-e122 [FREE Full text] [doi: 10.3414/ME17-05-0003] [Medline: 28925418]
35. Beratarrechea A, Moyano D, Irazola V, Rubinstein A. mHealth interventions to counter noncommunicable diseases in developing countries: still an uncertain promise. Cardiol Clin 2017 Feb;35(1):13-30. [doi: 10.1016/j.ccl.2016.08.009] [Medline: 27886783]
36. Crengle S, Smylie J, Kelaher M, Lambert M, Reid S, Luke J, et al. Cardiovascular disease medication health literacy among Indigenous peoples: design and protocol of an intervention trial in indigenous primary care services. BMC Public Health 2014 Jul 12;14:714 [FREE Full text] [doi: 10.1186/1471-2458-14-714] [Medline: 25016481]
37. Banbury A, Roots A, Nancarrow S. Rapid review of applications of e-health and remote monitoring for rural residents. Aust J Rural Health 2014 Oct;22(5):211-222. [doi: 10.1111/ajr.12127] [Medline: 25303412]

Abbreviations

BP: blood pressure
CVD: cardiovascular disease
HIC: high-income country
LMIC: low- and middle-income country
mHealth: mobile health
RCT: randomized controlled trial
SMS: short message service

Edited by G Eysenbach; submitted 04.04.19; peer-reviewed by E Ross, J Waite-Jones, E Meinert; comments to author 18.05.19; revised version received 12.07.19; accepted 18.08.19; published 07.10.19.

Please cite as:
Wali S, Hussain-Shamsy N, Ross H, Cafazzo J
Investigating the Use of Mobile Health Interventions in Vulnerable Populations for Cardiovascular Disease Management: Scoping Review
JMIR Mhealth Uhealth 2019;7(10):e14275
URL: https://mhealth.jmir.org/2019/10/e14275
doi:10.2196/14275
PMID:31593547

Review

Mobile Health for First Nations Populations: Systematic Review

Georgina R Hobson[1], BAppSc, MHlthProm; Liam J Caffery[1], BInfoTech, PhD; Maike Neuhaus[1], PhD, MPsych; Danette H Langbecker[1], BHlthSci, PhD

Centre for Online Health, Centre for Health Services Research, The University of Queensland, Woolloongabba, Australia

Corresponding Author:
Georgina R Hobson, BAppSc, MHlthProm
Centre for Online Health
Centre for Health Services Research
The University of Queensland
Ground Floor, Building 33
Princess Alexandra Hospital
Woolloongabba, QLD, 4102
Australia
Phone: 61 731761685
Email: g.hobson@uq.edu.au

Abstract

Background: The ubiquitous presence and functionality of mobile devices offers the potential for mobile health (mHealth) to create equitable health opportunities. While mHealth is used among First Nations populations to respond to health challenges, the characteristics, uptake, and effectiveness of these interventions are unclear.

Objective: This review aimed to identify the characteristics of mHealth interventions (eg, study locations, health topic, and modality) evaluated with First Nations populations and to summarize the outcomes reported for intervention use, user perspectives including cultural responsiveness, and clinical effectiveness. In addition, the review sought to identify the presence of First Nations expertise in the design and evaluation of mHealth interventions with First Nations populations.

Methods: The methods of this systematic review were detailed in a registered protocol with the International Prospective Register of Systematic Reviews (PROSPERO, CRD42019123276). Systematic searches of peer-reviewed, scientific papers were conducted across 7 databases in October 2018. Eligible studies had a primary focus on mHealth interventions with experimental or quasi-experimental design to respond to a health challenge with First Nations people from Canada, Australia, New Zealand, and the United States. Two authors independently screened records for eligibility and assessed risk of bias using the Joanna Briggs Institute checklists. Data were synthesized narratively owing to the mix of study designs, interventions, and outcomes. The review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.

Results: Searches yielded 1053 unique records, after review and screening, 13 studies (5 randomized controlled trials and 8 quasi-experimental designs) were included in the final analysis. Studies were conducted in Australia (n=9), the United States (n=2), and New Zealand (n=2). The most common health challenge addressed was mental health and suicide (n=5). Intervention modalities included text messaging (n=5), apps (n=4), multimedia messaging (n=1), tablet software (n=1), or a combination of short messaging service (SMS) and apps (n=1). Results showed mixed engagement with the intervention (n=3); favorable user perspectives, including acceptability and cultural appropriateness (n=6); and mixed outcomes for clinical effectiveness (n=10). A diverse range of risks of bias were identified, the most common of which included a lack of clarity about allocation and blinding protocols and group treatment for randomized controlled trials and a lack of control group and single outcome measures for quasi-experimental designs. First Nations expertise informed all mHealth studies, through authorship (n=8), affiliation with First Nations bodies (n=3), participatory study design (n=5), First Nations reference groups (n=5), or a combination of these.

Conclusions: mHealth modalities, including SMS and apps, appear favorable for delivery of health interventions with First Nations populations, particularly in the area of mental health and suicide prevention. Importantly, First Nations expertise was strongly embedded within the studies, augmenting favorable use and user engagement. However, evidence of efficacy is limited.

(JMIR Mhealth Uhealth 2019;7(10):e14877) doi:10.2196/14877

KEYWORDS

mHealth; mobile health; indigenous; First Nations; aboriginal; humans; systematic review

Introduction

Mobile phones and other wireless devices have the potential to disrupt traditional health service delivery by enabling consumers to engage with health information, comanage conditions, and gain support for health challenges [1-3]. This transformation, known as mobile health (mHealth), is underpinned by the near-ubiquitous presence and functionality of mobile devices, placing health-promoting initiatives in the hands of consumers [4,5]. Furthermore, mHealth may extend equitable health access to underserved populations, such as First Nations Peoples [6,7]. Research has begun to investigate the potential role of mHealth in responding to the health disparities which continue for First Nations people [8-13]. Jones et al report that alongside other technologies, mHealth is being adopted and adapted for cultural relevance within First Nations communities where meaningful participatory approaches are applied [14,15]. Enthusiasm and engagement with the use of mobile devices in health interventions with a variety of health challenges such as mental health and hepatitis B have been reported [14,16]. However, some concerns with regard to accessibility have been noted. Primarily, a digital divide exists through lack of reliable access to hardware, connectivity services and infrastructure depending on location [14,17], and mobile devices are often a shared resource within the community, resulting in intermittent access for some individuals [18,19].

To date, few reviews with a focus on mHealth with First Nations populations have been published. A scoping review by Brusse et al [15] identified peer-reviewed evidence within both a global, and Australia's First Nations context, for social media and mobile apps as health promoting modalities for smoking cessation, otitis media, and sexual health. Jones et al [14] conducted a critical review using Indigenous research methodology on the development and use of assistive technologies (ie, assistive devices, mHealth, and eHealth) in First Nations populations. These reviews provide valuable insights into the importance of user development, the paradox of a digital divide yet the high adoption of mobile technology, and use of social media for health interventions. However, there remains a need to synthesize the characteristics, quality, and outcomes of peer-reviewed experimental mHealth interventions with First Nations populations, across a wider range of health topics. Consideration of how mHealth research has engaged First Nations expertise is also needed to inform future research aiming to deliver health information and support to First Nations peoples.

The aim of this review was to systematically review the evidence for the use of mHealth interventions to address health challenges with First Nations populations. Specifically, the review aimed to identify characteristics of mHealth interventions (eg, study locations, health topic, and modality) evaluated with First Nations populations and to summarize the outcomes reported in relation to intervention use, user perspectives including cultural responsiveness, and clinical effectiveness. In addition, the review sought to identify the presence of First Nations expertise in the design and delivery of mHealth interventions with First Nations populations.

Methods

A systematic review of the peer-reviewed literature was conducted. A protocol outlining the methods of this systematic review were registered with the International Prospective Register of Systematic Reviews (PROSPERO, CRD42019123276) and published on their website [20]. The review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [21].

Eligibility Criteria

Eligible studies were experimental or quasi-experimental study designs using quantitative, qualitative, or mixed methods that focused on the use of mHealth to address a health challenge with First Nations populations. For the purpose of this review, mHealth was defined as per the World Health Organization [3], as the delivery of health interventions via wireless devices (ie, mobile phone, smartphone, cell phone, tablet, personal digital assistant, or pocket personal computer) using wireless modalities (ie, app, short messaging service (SMS), multimedia messaging service (MMS), voice messages, email, and program or website designed for mobile or tablet use). Studies were excluded if they involved mHealth interventions delivered via websites designed for nonmobile devices, interventions with specific equipment (eg, telehealth systems), and intelligent assistive technology using cognitive devices, physiologic or environmental sensors, and tracking capabilities. Studies reporting at least 1 of the following outcomes were included: (1) use of the mHealth intervention (eg, clicks, SMS replies, and logins); (2) user perspectives of the mHealth intervention (eg, acceptability, barriers, or enablers to its use, functionality, and cultural responsiveness); and (3) clinical effectiveness (eg, impact on screening, clinic attendance, medication compliance, and knowledge). Eligible papers were full text, published in English and reported outcomes of studies conducted in Australia, Canada, New Zealand, and the United States. These countries were the focus of this review given the similarities in colonialist history and policies with intergenerational impacts for First Nations people [14,22]. Studies were included if participants from these 4 countries were either First Nations (or if part of a multicultural sample, First Nations results were specifically reported) or parents or service providers (either indigenous or nonindigenous) of First Nations people. Research protocols, editorials, abstracts, reviews, and case reports were excluded.

Search Strategy

In October 2018, the following databases were searched systematically and without any limits to publication date: Cochrane Library, EMBASE, CINAHL, MEDLINE via PubMed, Scopus, Web of Science, and PsycINFO. A comprehensive list of search terms and strings for the key themes of *mHealth* and *First Nations or Indigenous* was developed with librarian assistance. Search strings included proximity operators, truncation, and phrase searching to explore possible iterations of both themes. To ensure capture of all relevant materials within the Australian Indigenous context, 3 additional Australia-specific searches were conducted in PubMed, CINAHL, and PsycINFO following recommendations

of the Lowitja Institute [23]. A sample of search strings is available in Multimedia Appendix 1.

Study Selection

Selection of relevant scientific papers followed the PRISMA flowchart [21]. Citations and abstracts were exported, and duplicates were removed. Screening of both the title and abstract and then remaining full-text papers was conducted independently by 2 of 3 authors at each stage (GH, DL, and MN), with discussions to reach consensus. Reasons for exclusion of full-text papers were noted. Manual searches of reference lists were conducted to identify any further studies.

Data Extraction and Risk of Bias

Data extracted from each full-text paper included the following: title, publication year, authors, country and region, conflicts of interest, database, study aims, intervention name, health challenge addressed, study design, setting, intervention focus, intervention modality, intervention content, control, key outcome measures and instruments, outcome assessment timing, outcome assessment modality, participants, recruitment method, response rate, loss to follow-up, and other notes. Intervention outcomes were noted under 3 main headings: use, user perspectives including perceived cultural responsiveness, and clinical effectiveness.

The extent to which First Nations expertise was present within or sought by research teams in the studies was assessed via data drawn from full-text papers using authorship lists or public author profiles online. Authors' statements regarding their identity or heritage were noted. Where such statements were unavailable, First Nations authorship was noted as *unknown*. Affiliations with First Nations bodies were recorded from full-text papers and Web-based profiles where available. Study designs and methods were reviewed to identify the use of participatory design principles or involvement of First Nations stakeholders in the intervention development or study conduct.

Risk of bias was assessed for each study using the relevant Joanna Briggs Institute (JBI) Critical Appraisal Checklist. This checklist assesses the methodological quality of a study and the extent to which the possibility of bias has been addressed in study design, conduct, and analysis [24,25]. Two authors (GH and LC) independently rated the risk of a range of biases for each included study as *yes*, *no*, *unclear*, or *not applicable* and tallied the number of *yes* responses, with consensus achieved via discussion. To enable comparison across studies using checklists with different numbers of items (9 items for quasi-experimental studies and 13 items for randomized controlled trials), a percentage was calculated. Level of Evidence (LoE) for effectiveness was also determined via a JBI tool that provides a rank of study designs "based on the likely best available evidence" [25] (p. 4). Study designs were compared with the 5 overall levels (*Level 1: Experimental* through to *Level 5: Expert Opinion and Bench Research*), with the specific level selected from possible subsets.

Data Synthesis

A narrative approach was used for data synthesis because of the mix of study designs and research approaches. Data were summarized under 7 headings as follows: (1) study characteristics, (2) design and content of mHealth Interventions, (3) risk of bias, (4) First Nations expertise, (5) intervention use, (6) user perspectives, and (7) clinical effectiveness .

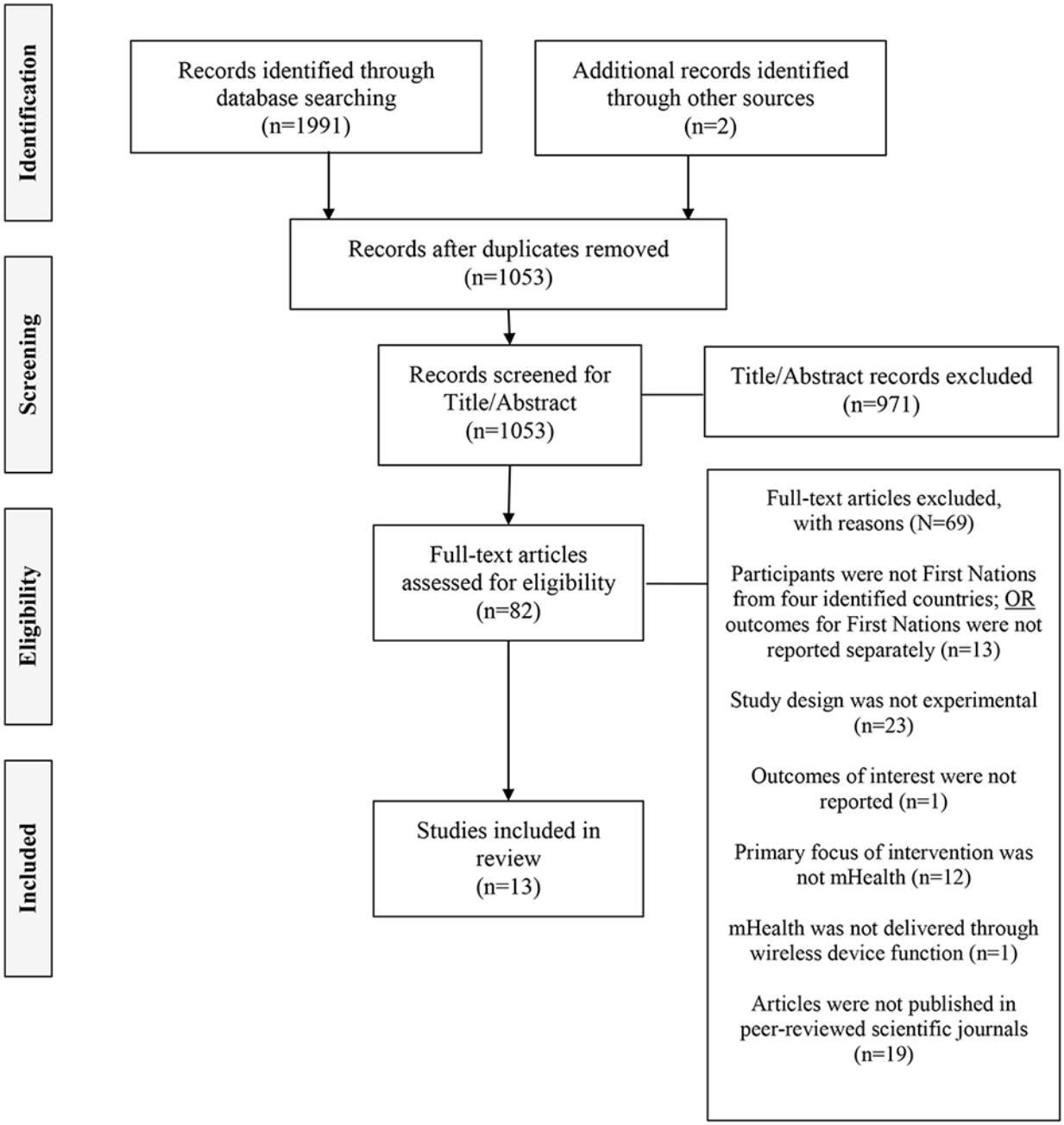
Results

From database searches, 1051 unique records were identified; 2 additional papers were located by manual searches, with a total of 1053 papers screened at the title and abstract level. After exclusion of 971 papers, 82 full-text papers were screened. A further 69 papers were excluded, most commonly due to a nonexperimental design (n=23) or absence of full text (n=19); 13 studies were included in the narrative synthesis (Figure 1).

Study Characteristics

All 13 studies included in this review were published between 2005 and 2018 (Multimedia Appendix 2): 5 studies were randomized controlled trials, where control groups were waitlisted for intervention (n=1) [26], received usual care (n=1) [27], had limited contact via SMS (n=2) [28,29], or received a combination of SMS, a paper resource, and usual care (n=1) [30]; 8 studies were single-arm quasi-experimental designs with a mix of qualitative and quantitative outcomes. Most studies were conducted in Australia (n=9), others in New Zealand (n=2) and the United States (n=2). No eligible studies were identified from Canada. All mHealth interventions were carried out in community settings or a combination of community and clinic and hospital settings. Participant demographics across the studies varied owing to the nature of the interventions. Adolescents and younger adults (15-30 years) were participants in 3 interventions focused on sexual health [31], new fatherhood [32], and suicide [26], whereas middle to older adults (40-75 years) were participants in studies relating to colorectal screening [27] and heart failure patient support [33]. In total, 2 studies recruited parents of younger children for interventions around child health [28,34], while other studies sought a wider sample of age for the intervention [29,30] or did not state participants' age (n=2) [35,36]; 8 studies included a higher proportion of female participants, while 2 studies had more males and 3 did not report gender. Most studies identified participants as either First Nations (n=6) [27,31-33,36,37] or service providers to First Nations clients, or parents and carers of First Nations children (n=4) [28,34,35,38]. In studies where participants were not solely First Nations (n=3), specific outcomes were reported for First Nations people [26,29,30].

Figure 1. Flowchart of identification, screening, eligibility, and inclusion of scientific papers. mHealth: mobile health.



Design and Content of mHealth Interventions

A wide range of mHealth designs, content, and intensities were used in the included studies (Multimedia Appendix 2). A total of 5 studies addressed mental health and suicide, while single studies focused on postpartum blood glucose screening, smoking, heart failure, colorectal cancer screening, chronic otitis media, hazardous drinking, infant feeding, and sexual health. Interventions were aimed at risk reduction, prevention or screening (n=7) [26,27,29-32,34], treatment and disease management (n=3) [28,33,36], or targeted multiple disease phases (n=3) [35,37,38]. The most common intervention modality was SMS (n=5) [27,29-32], followed by mobile apps (n=4) [26,35,37,38], MMS and SMS (n=1) [28], and via tablet software (not an app) (n=1) [33]. Two interventions used a combination of SMS and mobile phone or tablet apps [34,36].

Interventions Using Messaging (Short Message Service and Multimedia Message Service)

A total of 3 studies combined educational and supportive content via SMS to participants who were new fathers [32], wanted to quit smoking [29], or reduce their hazardous drinking levels [30]. Studies used SMS to prompt participants to attend medical clinics for colorectal cancer screening [27] or to promote sexually healthy behaviors with young Native Americans and American Indians [31]. A combination of education content and prompts for clinic appointments via SMS and MMS was used in 1 study to address chronic otitis media in children [28].

Interventions Using Apps

The *AIMhi Stay Strong* and *iBobbly* contained visual representations, voiceovers, action-based content, and goal

setting to build mental health and reduce risk of suicide. These 2 apps were used in interventions with service providers and community members in 4 studies either separately or in combination [26,35,37,38]. One study used a tablet-delivered program for education with heart failure patients [33].

Other Interventions

In Australia, infant feeding knowledge was shared with parents of Aboriginal children, primarily via app or a website and SMS version for parents without the app [34]. Another study to enhance blood glucose testing following diabetes in pregnancy used several contact methods with Aboriginal women including SMS, social media, email, mobile phone calls, and face-to-face meetings [36].

Across studies, intervention modality was not closely associated with the intensity of participant contact. In the early phase of a smoking intervention using SMS, participants were contacted up to 5 times per day [29], whereas a study using SMS prompts for cancer screening contacted participants 3 times (or less) over a 3-month period [27]. Interventions using apps were generally self-guided and ranged from use during a single session [37,38] to use in multiple sessions over a longer period. For example, participants in the *iBobbly* app trial were encouraged to regularly use the app over 6 weeks to progress through modules and self-assessments.

Risk of Bias

Using the JBI tool, LoE of the 13 studies were determined as either Level 1.c—Experimental, randomized controlled trial or Level 2.d—Quasi-experimental, Pretest—Posttest or retrospective control group study [25]. There was variability in the quality of the 5 randomized controlled trials (JBI Level 1.c), with a median quality score of 7.0 and a large range (6-11) out of 13 criteria to address risk of bias (Table 1). The 8 quasi-experimental studies (JBI Level 2.d) had a median quality score of 3.0 (range 2-6) out of 9 criteria for managing risk of bias. Once converted to percentages to allow comparison across all study designs, the risk of bias score varied across studies from 22% through to 85%. The randomized controlled trials met a higher number of quality criteria, and thus exhibited a higher mean percentage of 62%, whereas the quasi-experimental studies met fewer criteria with a mean percentage of 34%. The most common risks of bias identified in the randomized controlled trials were lack of clarity in allocation and or blinding protocols (n=5) [26-30] and absence of reliable outcome measures (n=3) [26,28,29]. Risk of bias commonly identified in the quasi-experimental studies were absence of control group (n=8) [31-38] and lack of clarity about treatment or care other than the intervention (n=7) [32-38].

A majority of the included studies (n=8) were subject to loss to follow-up. Studies acknowledged these modest (5%-12%) [26,27,29,33] or greater losses to follow-up (19%-49%) [28,30,31,36] and where appropriate, the losses to follow-up were largely accounted for through intention-to-treat analysis.

First Nations Expertise

A notable level of First Nations expertise was embedded in the mHealth studies through authorship, consultation, and research methodology (Table 1). A variety of resources were used to identify First Nations authorship including the following: full-text research papers, author profiles, or other Web-based sources [39-50]. The majority of studies (n=8), included at least 1 author who identified as First Nations in the country in which the study was conducted [27,29,30,32,33,36-38]. A total of 3 research teams based in Australia [26,34] and the United States [31] had close affiliations with First Nations bodies, some of which served as a source for participant recruitment. First Nations expertise was also sought through participatory design principles in 5 research studies [28,31-33,35] or through key reference groups composed of First Nations stakeholders (n=5) [26,27,33,35,37].

Cultural expertise was used to ensure appropriate language, imagery, and literacy from health interventions. In 3 studies, the importance of images, design, and audio were highlighted to ensure authentic representation of First Nations identity and assist users who face literacy challenges [26,33,38]. For example, participants using a mental health app highlighted the relevance of imagery given the First Nations significance of art for knowledge transfer throughout generations [35]. Language was a central reference point in 8 studies with consideration of word content for literacy key words, cultural relevance, and local dialects [27-33,38]. Consultation with First Nations stakeholders regarding cultural relevancy helped tailor SMS content for young Aboriginal fathers in Australia [32] but did not result in local language use for SMS prompts for colorectal cancer screening in Alaska [27]. Content of mHealth interventions which drew on cultural assets such as family and relationships, and a strength-based approach to First Nations health were noted in 3 studies [29,35,38], including the *AIMhi Stay Strong App* [38].

Intervention Use

Three studies reported participant use of the mHealth interventions with mixed results (Table 2) [26,32,36]. In a randomized controlled trial of a self-guided mental health app (*iBobbly*), 85% of participants with data available completed all 6 activities during the 6-week trial; however, use data was only available for 66% of trial participants [26]. A quasi-experimental study investigating the use of SMS to support young Aboriginal fathers in Australia also reported use by the majority of participants, with more than half clicking the link provided in 1 particular SMS (56%). The overall response rate to mood-tracking messages was 75% (106/141) [32]. In contrast, a quasi-experimental study [36] reported low response rates to clinic appointment prompts sent by SMS, phone call, email, and social media via mobile app. Although participants had nominated mobile phones as their preferred contact method, only 14% of contact attempts through the various mHealth modalities (including only 1 of 29 SMS) received a response.

Table 1. First Nations expertise within mobile health studies.

Study design and Study	First Nations authorship	Affiliation with First Nations bodies	Participatory design principles	First Nations reference group
Randomized controlled trial				
Bramley et al [29], smoking cessation	Yes	Unknown	Unknown	Unknown
Muller et al [27], colorectal cancer screening	Yes	Unknown	Unknown	Yes
Phillips et al [28], chronic otitis media	Unknown	Unknown	Yes	Unknown
Sharpe et al [30], hazardous drinking levels	Yes	Unknown	Unknown	Unknown
Tighe et al [26], suicide risk	Unknown	Yes	Unknown	Yes
Single-arm pre-post quantitative				
Yao et al [31], sexual health	Unknown	Yes	Yes	Unknown
Single-arm pre-post mixed				
Clarke et al [33], heart failure	Yes	Unknown	Yes	Yes
Dingwall et al [38], mental health	Yes	Unknown	Unknown	Unknown
Single-arm postquantitative				
Kirkham et al [36], postpartum blood glucose testing	Yes	Unknown	Unknown	Unknown
Fletcher et al [32], new fatherhood	Yes	Unknown	Yes	Unknown
Single-arm postqualitative				
Dingwall et al [35], mental health	Unknown	Unknown	Yes	Yes
Houston et al [34], infant feeding	Unknown	Yes	Unknown	Unknown
Povey et al [37], mental health and suicide risk	Yes	Unknown	Unknown	Yes

Table 2. Intervention use reported across included studies.

Study design and study	Measures of intervention use; follow-up timing	Key findings
Randomized controlled trial		
Tighe et al [26], suicide risk	App activity completion (total of 6); during intervention over 6 wks[a]	Participant completion: 6 activities=85% (34/40); 5 activities=2% (1/40); 2 activities=13% (5/40)
Single-arm postquantitative		
Fletcher et al [32], new fatherhood	Click rates for SMS[b] links, response to SMS (mood trackers); during intervention over 6 wks	Participant click rates: 56% *Routines: Aboriginal and Torres Strait Islander Parents*, 41% *Baby Talk*, 4% *Crying*, 0% *Bonding for Dads* and *Postnatal depressionandwomen*. Participant response to mood trackers: 90% to at least 1×SMS, 75% to total SMS during trial (106/141)
Kirkham et al [36], postpartum blood glucose testing	Message response via mHealth (incl. SMS, phone call, email, and social media); during intervention over 24 m[c]	14% (18/252) total responses via mHealth, including 3% response to SMS (1/29)

[a]wk(s): week(s).

[b]SMS: short messaging service.

[c]m: month.

Table 3. User perspectives regarding the mobile health interventions.

Study design and study	Construct measures; follow-up timing	Key findings
Randomized controlled trial		
Phillips et al, [28] chronic otitis media and ear perforation, N=53	Willingness to receive future child health SMS[a] and MMS[b] and feedback (via survey) 6 wks[c] from baseline	Willingness: Yes=76% (37/49), no significant difference between intervention and control groups, preference for SMS clinic prompt versus MMS=33% (3/9). Other feedback: overall interest and appreciation for message content, many shared videos
Single-arm pre-post mixed		
Clarke et al, [33] heart failure, N=5	Mean participant satisfaction score (via survey), other feedback (via interview), Single use (1 hr[d])	Satisfaction mean score=4.15/5. Other feedback: Acceptance: *I liked it all*, the teaching tool was *good*. Comprehension: *no big words*, *seeing that really did make me realize*. Impact: *This is something I will never forget*. Usability: some frustration with touchscreen (1/5)
Dingwall et al, [38] mental health, N=130	Participant feedback (via open-ended survey), single use (1 day)	Challenge to implementation: Technology availability (access to iPads and WiFi). Support for implementation: Practice
Single-arm postqualitative		
Dingwall et al, [35] mental health, N=15	Acceptability, feasibility, applicability of app (via interviews and thematic analysis), 1 m[e]	Positive responses: Acceptability: visual appeal, ease of use, cultural relevance, innovative format, etc. Building relationships: help therapeutic relationships, shift power imbalance, build client ownership, etc. Applicability: broad suitability for age and regions, mixed views concerning digital literacy. Constructive feedback: Constraints to implementation: technology availability, time, staff, local language. Integration with systems: data merges with existing records. Training recommended for content and processes
Houston et al, [34] infant feeding, N=10	Viability, cultural appropriateness, design and language, target audience of app (via interviews and thematic analysis), 6 wks	Positive responses to viability: Familiar technology and source of information, app better than website, language, and content-trustworthy, useful, helpful, consistent and reassuring, single source of info was valuable versus searching websites. Mixed views on usability: Some frustrations with app functionality, technical difficulties, and amount of content but did not prevent use of app. Mixed views on cultural appropriateness: Service providers felt more cultural responsiveness was needed; however, parents did not. Suggestions to improve relevance: integration of personal stories, content for other carers, and guidance from key community member. Mixed views on design: Additional images, colors and art work was desired
Povey et al, [37] mental health and suicide risk, N=9	Acceptability of *AIMhi Stay Strong App* and *iBobbly* assessed via interviews and thematic analysis, single use, same day	Overall enthusiasm and optimism for app concepts and progressive support for mental health. Influencers of acceptability: Person: illness, history, tech competence, literacy, and local language; Environment: community awareness, stigma, and support; App: content, graphics, animation, ease of use, access, security, and information sharing

[a]SMS: short messaging service.

[b]MMS: multimedia messaging service.

[c]wk(s): week(s).

[d]hr: hour.

[e]m: month.

User Perspectives

A total of 6 studies reported participants' views of acceptability, cultural safety, or relevance for mHealth to deliver health interventions (Table 3). In that, 1 study used phone messaging [28] and 5 studies involved apps (Table 3) [33-35,37,38]. User perspectives were reported using quantitative outcomes from surveys or interviews [28,33] and qualitative data from interviews and focus groups [34,35,37].

User perspectives on phone messaging: All participants of a randomized controlled trial reported SMS and MMS communications as culturally appropriate [28]. The majority (76%) indicated they would be happy to receive health messages in the future, although 3 of 9 participants preferred SMS clinic reminders to educational MMS [28].

User perspectives on tablet or phone apps: An intervention delivered via tablet software in a quasi-experimental study

employed a survey with a 5-point Likert scale to assess aspects of navigation and usability. Participants responded to positively phrased statements with 1=strongly disagree to 5=strongly agree, and where statements were phrased negatively, scores were reversed [33]. The mean participant satisfaction score for the heart failure education program was calculated at 4.15/5 (n=5). Participants also answered open-ended questions relating to acceptance, comprehension, and impact of the app. User satisfaction scores and open-ended feedback indicated user comprehension, positive impact, and overall *acceptance* of the program and its modality, despite being the first occasion each participant had used a tablet [33].

The other 3 quasi-experimental studies reported qualitative data for user perspectives from interviews and focus groups. Mental health service providers for Aboriginal and Torres Strait Islander people expressed positive feedback for the *AIMhi Stay Strong App* through themes of acceptability, building relationships, and applicability [35]. Constructive feedback regarding barriers to practical use of the app were also recorded. Povey et al [37] gathered user perspective data on factors that would impact acceptability of the *AIMhi Stay Strong* and *iBobbly* apps. Focus group themes included characteristics of the person, the environment, and the app. Participants expressed enthusiasm for the progressive nature of the apps in supporting Australia's First Peoples with mental health challenges. Local language was referred to as a key element for acceptability and transferability [35,37,38].

Qualitative data from parents who used an infant feeding app indicated the intervention was well received and culturally appropriate, despite no cultural tailoring [34]. In contrast, service providers (staff) felt the lack of cultural consideration could be a barrier to use. However, parent participants viewed the app as a viable delivery method for quality content. Parent participants also expressed their familiarity with mobile phone technology and a preference for the app versus a website. Some technology frustrations were experienced by participants; however, this did not inhibit their use of the app. Parents suggested embedding personal stories in the app to further connect them with program content and lift cultural relevance [34].

Clinical Effectiveness

A total of 10 studies reported data relating to the impact of an mHealth intervention, although the outcomes varied significantly owing to the nature of interventions, study designs, and outcome measures (Table 4). Of the 10 studies, 6 were delivered using phone messaging (SMS and MMS), including 4 randomized controlled trials [27-30] and 2 quasi-experimental studies [31,36]. The remaining 4 studies involved apps, tested through randomized controlled trial (n=1) [26] or quasi-experimental designs (n=3) [33,34,38].

Clinical Effectiveness for Phone Messaging

The 4 randomized controlled trials of phone messaging interventions assessed diverse health outcomes [27-30]; 3 studies reported statistically significant outcomes. Bramley et al [29] reported the relative risk of smoking cessation for participants who received an SMS intervention was 2.34 at 6 weeks (95% CI 1.44 to 3.79; *P*<.001) compared with controls. Self-report quit rates at 12 and 26 weeks remained high but were not statistically significant compared with controls [29]. Muller et al [27] reported that the hazard ratio of colorectal cancer screening was significantly higher after an SMS intervention compared with control (1.30 hazard ratio; 95% CI 1.04 to 1.62; *P*=.02). However, when stratified by age group, intervention participants' screening remained higher than control, but were not statistically significant [27]. Hazardous drinking scores were significantly improved following an SMS intervention compared with control participants at 3, 6, and 9 months in a study by Sharpe et al [30]. Following discharge from hospital, intervention participants reported significantly lower scores compared with controls on the Alcohol Use Disorders Identification Test—Consumption at 3 months (−0.322; 95% CI −0.636 to −0.008), 6 months (−0.296; 95% CI −0.474 to −0.118), and 9 months (−0.260; 95% CI −0.463 to −0.057) [30]. While Alcohol Use Disorders Identification Test—Consumption scores were not reported separately for participants, preplanned secondary analysis revealed that the intervention was effective regardless of ethnicity.

In contrast, Phillips et al [28] reported no significant difference for impact of an SMS and MMS intervention on the primary outcome of clinic attendance. No mean difference was recorded for clinic visits per child for any reason (−0.1; 95% CI −1.1 to 0.9; *P*=.90) or clinic visit per child for ear health (−0.3; 95% CI −0.8 to 0.2; *P*=.50). In addition, no significant differences were observed for the secondary outcome of diagnoses [28].

Two quasi-experimental studies involving phone messaging reported mixed clinical outcomes related to sexual health [31] and postpartum glucose testing [36]. Yao et al reported significant improvements from baseline to 1 week post-SMS for condom use and sexually transmitted infection (STI)/HIV screening for the following: (1) attitude to condom use (odds ratio [OR] 3.25; 95% CI 1.44 to 7.35; *P*=.005); (2) condom use behavior (OR 2.43; 95% CI 1.15 to 5.13; *P*=.02); and (3) intention to undergo STI/HIV testing (OR 2.46; 95% CI 1.42 to 4.26; *P*=.01). These positive outcomes were sustained at 3 months despite nonsignificant shifts in other sexual health constructs (knowledge, intention, and self-efficacy), and despite notable loss to follow-up in surveys [31]. Kirkham et al reported a lack of significant difference in completion of postpartum glucose testing for women (n=52) who were *responders* to clinic reminder messages (55%) versus those who were not (43%). Although some differences were observed in completion of certain tests for *responders* versus *nonresponders*, conclusive statements are impacted by loss to follow-up (n=10/52) and the small sample size. Furthermore, this study used other modes of communication than mHealth, and contact method was not reported against participation [36].

Table 4. Clinical effectiveness of mobile health interventions.

Study design, study, and sample size	Construct measures; follow-up timing	Key findings
Randomized controlled trials		
Bramley et al [29], smoking cessation, N=355[a]	RR[b] self-report smoking cessation at 6 wks[c], 12 wks, 26 wks	2.34 RR versus control (6 wk) *P*<.001, 1.37 RR versus control (12 wk) *P*=.11, 1.17 RR versus control (26 wk) *P*=.46
Sharpe et al [30], hazardous drinking, N=126	Mean difference in hazardous drinking scores (AUDIT-C)[d] at 3 m[e], 6 m, 12 m	−0.322 versus control (3 m) *P*=.04, −0.296 versus control (6 m) *P*=.002, −0.260 versus control (12 m) *P*=.002
Muller et al [27], colorectal cancer screening, N=2386	HR[f] of cancer screening in medical register at 6 m	1.30 HR versus control (all ages) *P*=.02, 1.42 HR versus control (50-75 years) *P*=.07, 1.24 HR versus control (40-49 years) *P*=.12
Phillips et al [28], clinic attendance and chronic otitis media, N=53	Mean difference in clinic attendance, diagnosis of chronic otitis media, or ear perforation at 6 wks	Clinic attendance: −0.1 mean difference clinic visits versus control *P*=.90, −0.3 mean difference clinic visit ear health versus control *P*=.50; Risk difference for diagnosis: no perforation 6% (−10, 20) *P*=.60, acute otitis media with perforation N/A, dry perforation −5% (−30, 20) *P*=.70, chronic suppurative otitis media −1% (−30, 30) *P*>.99
Tighe et al [26], suicide risk, N=61	Effect size (Cohen *d*) for suicidality (DSI-SS[g]), impulsivity (BIS[h]), depression (PHQ-9[i]), psychological distress (K-10[j]) at 6 wks	DSI-SS=0.00 (95% CI −0.51 to 0.51) *P*=.30[k], BIS=not reported, PHQ-9=0.71 (95% CI 0.17 to 1.23) *P*=.007, K-10=0.65 (95% CI 0.12 to 1.17) *P*=.02
Single-arm pre-post quantitative		
Yao et al [31], sexual health, N=408	OR[l] condom use at 1 wk and 3 m: (1) knowledge, (2) attitude, (3) intention—partner and self, (4) behavior; OR test STI[m]/HIV at 1 wk and 3 m: (5) self-efficacy make appt, (6) self-efficacy speak to HCP[n], (7) attitude, (8) intention	OR at 1 wk: (1) 1.15, *P*=.70; (2) 3.25, *P*=.005; (3) partner 1.33, *P*=.30; (3) self 0.97, *P*=.92; (4) 2.43, *P*=.02; (5) 1.63, *P*=.50; (6) 1.01, *P*=.99; (7) 1.11, *P*=.81; (8) 2.46, *P*<.001. OR at 3 m: (1) 1.30, *P*=.48, (2) 3.93, *P*=.002; (3) partner 1.07, *P*=.82; (3) self 0.87, *P*=.61; (4) 2.60, *P*=.02; (5) 6.77, *P*=.10; (6) 1.12, *P*=.75; (7) 0.56, *P*=.15; (8) 2.64, *P*<.001
Single-arm postquantitative		
Kirkham et al [36], postpartum blood glucose screening, N=52	Attendance postpartum glucose testing (random glucose, fasting glucose, OGTT[o] and/or HBA_{1C}[p]) for message responders. Follow-up timing unclear	55% (12/22) responders any test versus 43% (13/30) nonresponders (nonsignificant) 32% (7/22) responders for OGTT versus 7% (2/30) nonresponders (nonsignificant)
Single-arm pre-post mixed methods		
Clark et al [33], heart failure, N=5	Heart failure knowledge via questionnaire, self-care indicators (maintenance, confidence, and management) via SCHFI[q] at same day	Knowledge (/20): mean score +2.0. SCHFI indicators (/100): Maintenance: mean score +16.7 (SD 25.2), Confidence: mean score + 44.4 (SD 20.8), Self-management: mean score +1.0 (SD 18.2)
Dingwall et al [38], mental health, N=138	Pre-post knowledge and confidence (12 items) for *AIMhi Stay Strong App* via visual analog scales and open-ended questions, likelihood of using *AIMhi Stay Strong App* in next 6 m at same day	Total sample: Increase in mean scores across 10 items postintervention *P*≤.01. Subgroups: No difference between groups for mean scores across 10 items postintervention; significant difference between groups with First Nations participants' mean scores lower for *confidence in use of other computers*, *P*<.01. Intention to use app in next 6 m: 83% total sample
Single-arm postqualitative		
Houston et al [34], infant feeding, N=10	Self-report infant feeding knowledge and practice via interview at 6 wks	Self-report increases in infant feeding knowledge that informed feeding practices

[a]Number of First Nations participants from total sample.

[b]RR: relative risk.

[c]wk(s): week(s).

[d]AUDIT-C: Alcohol Use Disorders Identification Test—Consumption.

[e]m: month.

[f]HR: hazard ratio.

[g]DSI-SS: Depressive Symptom Inventory—Suicidality Subscale.

[h]BIS: Barratt Impulsivity Scale.

[i]PHQ-9: personal health questionnaire-9.

[j]K-10: Kessler-10.

[k]*P* value for the interaction of intervention arm by time (preintervention vs postintervention).

[l]OR: odds ratio.

[m]STI: sexually transmitted infection.

[n]HCP: health care professional.

[o]GTT: oral glucose tolerance test.

[p]HBA_{1c}: hemoglobin A_{1c}.

[q]SCHFI: self-care heart failure inventory.

Clinical Effectiveness for Mobile Apps

A randomized controlled trial reported clinical effectiveness measures for the self-guided *iBobbly* app which includes the following: suicidality (Depressive Symptom Inventory—Suicidality Subscale [DSI-SS]), impulsivity (Barratt Impulsivity Scale), depression (Personal Health Questionnaire), and psychological distress (Kessler-10). Tighe et al [26] reported no significant reduction in the primary outcome of suicidal ideation between intervention and waitlist groups at 6 weeks (mean DSI-SS scores=1.9 in both groups), despite a reduction in pre-post mean DSI-SS scores for the intervention arm (pre=2.7, SD 2.2; post=1.9, SD 2.1). Secondary outcomes for intervention versus waitlist groups at 6 weeks were mixed, with no significant change in impulsivity, yet significant improvements in scores for depression (Personal Health Questionnaire-9=8.9, SD 5.4 vs 12.8, SD 5.5) and psychological distress (Kessler-10=22.7, SD 7.4 vs 27.9, SD 8.0) [26].

Three quasi-experimental studies reported outcomes for participant knowledge following interventions using phone or tablet-based apps or programs. Houston et al [34] supported parents via an infant feeding app over 6 weeks. Reports included an increase in knowledge and alterations to their infant feeding practices. Service providers trained in use of a mental health app over 1 day by Dingwall et al showed significant increases in knowledge and confidence across 10 of 11 items assessed in a questionnaire with visual analogue scales [38]. Similarly, Clark et al [33] reported participants' mean increase of 2 points on heart failure knowledge scores out of a possible 20 points (9.6, SD 1.3 to 11.6, SD 1.9) following 1-h use of a tablet-based education program. Increases in mean Self-Care Heart Failure Index scores were also reported. However, the results should be interpreted with caution as they were accompanied by incomplete Self-Care Heart Failure Index surveys (n=2) and high standard deviations, reflecting the small sample (n=5).

Discussion

Principal Findings

Mobile Health Research Characteristics—Studies and mHealth Interventions

There is limited reporting on the characteristics of mHealth interventions evaluated with First Nations Peoples, with 13 studies identified in this review. A number of studies were randomized controlled trials involving messaging, commonly available across devices and platforms through SMS and MMS. Previous studies suggest text messaging is efficacious, relatively easy to use and a low-cost modality [2,3,15,51]. The messaging-based interventions in this review drew on these strong characteristics to respond to health challenges with First Nations populations. Other studies used digital add-ons, including unique mobile apps or programs using mobile software. Despite their prolific presence and adoption by consumers, research into the health impact of apps lags behind that of SMS. This reflects the more recent development of this modality and subsequent recognition of a public health potential [52].

Intervention content and contact intensity varied significantly owing to the wide range of health topics addressed with First Nations populations. However, mental health and suicide prevention, particularly in the Australian context, were common themes (n=5). This is an important finding given the documented mental health challenges experienced by First Peoples of Australia, New Zealand, and the United States on account of intergenerational trauma [22]. It demonstrates that scientific research is responding to the mental and emotional challenges faced by First Nations populations [53].

Furthermore, this review noted that the diversity of First Peoples was recognized through collaborative projects, designed and trialed in local contexts, with consideration of the transferability of mHealth products rather than a one-size-fits-all approach [16,37]. Research teams were embedded with First Nations knowledge and leadership to explore the potential role of technology for health. First Nations expertise strengthened the relevance and transfer of the mHealth research and influenced participant engagement and outcomes for the interventions.

Mobile Health Outcomes

The diversity of mHealth aims and interventions yielded a mix of reported key outcomes across 3 domains: *use*, *user perspectives*, and *clinical effectiveness*.

First, despite participant assurance of access to or ownership of mobile devices, engagement levels with messaging and apps were varied [26,32,36]. While only 3 studies reported on this outcome, it is feasible that cultural tailoring and First Nations expertise during mHealth development enhanced the use of mHealth in 2 trials [26,32]. Authors of the study with lower use [36] acknowledged the wider social and environmental priorities that may influence engagement with an intervention. In particular, mobile devices may be viewed as a shared resource within First Nations communities [18,19,36], thus impacting the potential exposure of an mHealth intervention. Furthermore, the accuracy of data relating to phone ownership and use in First Nations communities (especially remote areas) is

problematic given a digital divide shaped by location, network infrastructure, cost of devices, and network access [17,52].

Second, overall favorable user perspectives were reported for 6 Australian mHealth studies [28,33-35,37,38]. High acceptability and user satisfaction were noted, even for participants without previous experience with devices [33,37]. Navigation difficulties and technical challenges with apps were a concern for some [37], although other participants were not deterred [34]. Notably, First Nations participants reflected on the progressive nature of mHealth to support communities in novel ways [35,37]. Others resonated with emphasis on First Nations assets (family and relationships) in the programs [35] and expressed the likelihood of using an mHealth service in the future if available [28,33,38]. Interestingly, some concerns about cultural safety by service providers were not mirrored by service recipients [34,35,37,38]. The acceptability of mHealth interventions reported here is in line with descriptions by Brusse et al [15], who noted high penetration and rapid uptake of mobile interfaces such as social media, apps, and messaging. Importantly, beyond social or entertainment value, mobile devices offer significant potential for access to health support in scenarios where individuals hold concerns about face-to-face interaction. For example, tablets may facilitate help-seeking behavior for young people who experience concerns about social interaction and moral judgement when accessing mental health support [26,37,38,53,54]. This systematic review identified significant First Nations expertise that conceivably augmented favorable user perspectives of the mHealth interventions. For example, the *AIMhi Stay Strong App* has been subject to significant work on cultural feasibility, design, and usability with guidance from First Nations stakeholders, and has been met with highly positive feedback by service providers [35,37,38]. Genuine cultural consultation through collaborative research and participatory design in other studies has been associated with high consumer interest [14] and considered paramount for cultural relevance and safety [16,51].

The clinical effectiveness of the mHealth interventions with First Nations populations in this review, while mixed, were in agreement with other reports. Although considered ideal for health interventions [4,6,7], conclusive statements for the clinical efficacy of mHealth are limited. In a systematic review of systematic reviews (n=23) for mHealth research, Marcolino et al [2] reported that evidence for mHealth efficacy was limited to several chronic health interventions including asthma and smoking cessation, mostly via basic messaging functionality of phones. Authors emphasized that a majority of studies in the reviews were of low methodological quality. Many of the included reviews found either no significance or conflicting outcomes, with a need for further studies to address quality and validation of pilot work with repeat trials of longer duration. The present review included a high heterogeneity across study designs, interventions, risk of bias, and outcome measures, limiting conclusive statements around effectiveness or efficacy.

The challenges of mHealth research described by Marcolino et al [2] and the present systematic review correspond with broader public health intervention research, where evidence for causality is often constrained by sample size, time, ethical considerations, and costs. Although these limitations can result in lower levels of empirical evidence, health-promoting interventions should also be evaluated on their "completeness and transferability" (p. 125) [55] to allow authentic evidence within the complex environment in which real-world interventions are trialed. Ben-Zeev et al emphasized the challenges associated with development and trial of mHealth interventions for use "in the wild" (p. 158) [56], noting that they require realism, collaboration, and flexibility of research teams. Several recent app trials have demonstrated the research possibilities and broader significance of mHealth interventions beyond solely clinical data, with encouraging results. For example, while research reports in this review [35,37,38] for the *AIMhi Stay Strong App* would traditionally be considered low levels of evidence [55], this app is informed by research showing that motivational care planning (MCP) effectively improved measures of well-being, substance misuse, and self-management among Aboriginal and Torres Strait Islander clients with chronic mental illness [57]. The subsequent adaptation of MCP into the *AIMhi Stay Strong App* could be deemed an mHealth success story as it is now used by health and community services in Australia [58]. This app has undergone several reiterations over recent years to suit a variety of population groups and organizational needs (personal communication, J Povey, May 14, 2019). It is also currently being trialed as a well-being intervention with Aboriginal and Torres Strait Islander people with chronic kidney disease [59]. In addition, while the pilot trial of the *iBobbly* app did not find clinically significant results for suicide ideation, the significant outcomes for depression and psychological distress have justified further app development and a national randomized controlled trial to continue the important work in this area [53].

Limitations

The ever-evolving nature and terminology of mobile health technology presents some challenges. To ensure consistent inclusion and exclusion of intervention studies, this paper adhered to the World Health Organization definition of mHealth [3], which includes text messaging, mobile apps, and other wireless modalities. Although this definition facilitates inclusion of a wide range of interventions, thus strengthening the review, it is acknowledged that some limits could have been set by the authors in an effort to manage the breadth of mHealth interventions. Publication bias may also have impacted the identification of mHealth studies with First Nations populations, whereby studies with inconclusive or negative findings may not have been published [60]. In addition, this review did not include unpublished (*gray*) literature, which may have yielded further studies examining mHealth interventions. Quality assessment was conducted using 2 JBI tools to allow risk of bias comparison across studies using differing designs, with 2 authors independently assessing each study to minimize error in the quality assessment. However, it is recognized that quality assessments may reflect limitations in reporting (eg, owing to publication space) rather than study conduct. Similarly, First Nations expertise was assessed using desktop research only. Information on authorship, key reference groups, participatory design, and affiliations was not always clear or available. The authors acknowledge that studies may have drawn on additional features of First Nations expertise.

Conclusion

This is the first systematic review to document the characteristics, use, user perspectives, and clinical effectiveness of mHealth from peer-review scientific studies, with unrestricted health topics, where participants are First Nations people, or parents or service providers to First Peoples from Australia, New Zealand, or the United States. Encouragingly, mHealth interventions have been conducted with considerable First Nations expertise, rather than mere consultation, to guide cultural usability and meaning for community. This is a meaningful finding given that First Nations knowledge is the most valuable asset available to address challenges of First Nations health and community [61]. It is a feasible conclusion that embedding First Nations expertise into the projects helped augment favorable reports and feedback for use and user perspectives in the mHealth interventions. Difficulty remains, however, in the ability to draw conclusive statements about clinical effectiveness given the heterogeneity of mHealth interventions with First Nations populations. Clinically significant outcomes have been reported for both messaging and app modalities. While messaging appears to have a more mature evidence base, apps are acceptable, and they are being used effectively with First Nations Peoples. Trials are underway to validate previous findings of significance and to explore the transferability of mHealth across the diversity of First Nations populations. Importantly, this ongoing research, particularly in the area of mental health and suicide prevention, holds potential for significant impact with First Nations communities.

Acknowledgments

We acknowledge the funding and support provided for this research by the Centre for Online Health and the Australian Government—Department of Health (Dementia and Aged Care Services). We also wish to acknowledge Ms Julie Hansen and Scott Macintyre from The University of Queensland Library services who provided advice and direction for the database searches. The authors have endeavored to use the terminology provided in The Lowitja Institute Style Guide (version 14) throughout this systematic review [62].

Authors' Contributions

GH, DL, and MN conceptualized the study and designed the review with input from LC. GH, DL, and MN designed the search strategies. GH performed database searches and GH, DL, and MN performed independent screening of abstract and full-text papers. GH completed data extraction and GH and LC conducted independent risk of bias assessments. GH wrote drafts of the review paper with significant input from DL and additional input from LC and MN. All authors contributed to the final draft and have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample search strategies, October 2018.
[PDF File (Adobe PDF File)90 KB - mhealth_v7i10e14877_app1.pdf]

Multimedia Appendix 2
Study design, risk of bias, and mobile health (mHealth) intervention characteristics.
[PDF File (Adobe PDF File)191 KB - mhealth_v7i10e14877_app2.pdf]

References

1. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth assessment: conceptualization of a global framework. JMIR Mhealth Uhealth 2017 May 2;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
2. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
3. World Health Organization. Global Diffusion of eHealth - Making Universal Health Coverage Achievable: Report of the Third Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2016.
4. Brown EL, Ruggiano N, Li J, Clarke PJ, Kay ES, Hristidis V. Smartphone-based health technologies for dementia care: opportunities, challenges, and current practices. J Appl Gerontol 2019 Jan;38(1):73-91. [doi: 10.1177/0733464817723088] [Medline: 28774215]
5. 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]

6. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
7. Chou WY, Prestin A, Lyons C, Wen KY. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
8. Taylor K, Guerin PT. Health Care and Indigenous Australians: Cultural Safety in Practice. Second Edition. South Yarra: Palgrave Macmillan; 2014.
9. Paradies Y. Racism. In: Carson B, Dunbar T, Chenhall RD, Bailie R, editors. Social Determinants of Indigenous Health. Crows Nest: Allen & Unwin; 2007:65-86.
10. Australian Indigenous HealthInfoNet. 2018. Overview of Aboriginal and Torres Strait Islander health status 2017 URL: https://healthinfonet.ecu.edu.au/healthinfonet/getContent.php?linkid=609266&title=Overview+of+Aboriginal+and+Torres+Strait+Islander+health+status%2C+2017 [accessed 2019-05-22] [WebCite Cache ID 78D2uXxXO]
11. Canadian - Government of Canada Publications. 2014. A Statistical Profile on the Health of First Nations in Canada: Determinants of Health, 2006 to 2010 URL: http://publications.gc.ca/site/archivee-archived.html?url=http://publications.gc.ca/collections/collection_2014/sc-hc/H34-193-1-2014-eng.pdf [accessed 2019-05-14] [WebCite Cache ID 78M6TW0JN]
12. Carroll M, Cullen T, Ferguson S, Hogge N, Horton M, Kokesh J. Innovation in Indian healthcare: using health information technology to achieve health equity for American Indian and Alaska native populations. Perspect Health Inf Manag 2011 Jan 1;8:1d [FREE Full text] [Medline: 21307987]
13. Tatau Kahukura Māori Health Chart Book 2015. Third Edition. Wellington, New Zealand: Ministry of Health; 2015.
14. Jones L, Jacklin K, O'Connell ME. Development and use of health-related technologies in indigenous communities: critical review. J Med Internet Res 2017 Jul 20;19(7):e256 [FREE Full text] [doi: 10.2196/jmir.7520] [Medline: 28729237]
15. Brusse C, Gardner K, McAullay D, Dowden M. Social media and mobile apps for health promotion in Australian indigenous populations: scoping review. J Med Internet Res 2014 Dec 10;16(12):e280 [FREE Full text] [doi: 10.2196/jmir.3614] [Medline: 25498835]
16. Davies J, Bukulatjpi S, Sharma S, Caldwell L, Johnston V, Davis JS. Development of a culturally appropriate bilingual electronic app about Hepatitis B for indigenous Australians: towards shared understandings. JMIR Res Protoc 2015 Jun 10;4(2):e70 [FREE Full text] [doi: 10.2196/resprot.4216] [Medline: 26063474]
17. Radoll P, Hunter B. Centre for Aboriginal Economic Policy Research - ANU. 2017. Dynamics of the Digital Divide URL: https://caepr.cass.anu.edu.au/research/publications/dynamics-digital-divide [accessed 2019-05-14] [WebCite Cache ID 78M8yyL4T]
18. Auld G, Snyder I, Henderson M. Using mobile phones as placed resources for literacy learning in a remote indigenous community in Australia. Lang Educ 2012 Jul;26(4):279-296. [doi: 10.1080/09500782.2012.691512]
19. Tangentyere Council, Alice Springs. 2007. Ingerrekenhe Antirrkweme: Mobile Phone Use Among Low Income Aboriginal People: A Central Australian Snapshot URL: https://www.tangentyere.org.au/publications/research_reports/MobilePhone.pdf [accessed 2019-05-14] [WebCite Cache ID 78M9SPNsY]
20. Hobson G, Neuhaus M, Caffery L, Langbecker D. Prospero - University of York. 2019. Evidence for the Use of mHealth Interventions to Address Health Challenges With Indigenous Populations: A Systematic Literature Review URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=123276 [accessed 2019-05-14] [WebCite Cache ID 78M6nRkNF]
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
22. Pulver LJ, Haswell MR, Ring I, Waldon J, Clark W, Whetung V, et al. World Health Organization. 2010. Indigenous Health – Australia, Canada, Aotearoa New Zealand and the United States - Laying Claim To a Future That Embraces Health for Us All URL: https://www.who.int/healthsystems/topics/financing/healthreport/IHNo33.pdf [accessed 2019-05-08] [WebCite Cache ID 78CslRluc]
23. Tieman JJ, Lawrence MA, Damarell RA, Sladek RM, Nikolof A. LIt.search: fast tracking access to Aboriginal and Torres Strait Islander health literature. Aust Health Rev 2014 Nov;38(5):541-545. [doi: 10.1071/AH14019] [Medline: 25109618]
24. Joanna Briggs Institute. 2018. Critical Appraisal Tools URL: http://joannabriggs.org/research/critical-appraisal-tools.html [accessed 2019-05-14] [WebCite Cache ID 78M7Alha7]
25. Joanna Briggs Institute. 2014. Supporting Document for the Joanna Briggs Institute Levels of Evidence and Grades of Recommendation URL: https://joannabriggs.org/sites/default/files/2019-05/JBI%20Levels%20of%20Evidence%20Supporting%20Documents-v2.pdf [accessed 2019-08-12] [WebCite Cache ID 78M89wdQA]
26. Tighe J, Shand F, Ridani R, Mackinnon A, de la Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]

27. Muller CJ, Robinson RF, Smith JJ, Jernigan MA, Hiratsuka V, Dillard DA, et al. Text message reminders increased colorectal cancer screening in a randomized trial with Alaska native and American Indian people. Cancer 2017 Apr 15;123(8):1382-1389 [FREE Full text] [doi: 10.1002/cncr.30499] [Medline: 28001304]
28. Phillips JH, Wigger C, Beissbarth J, McCallum GB, Leach A, Morris PS. Can mobile phone multimedia messages and text messages improve clinic attendance for Aboriginal children with chronic otitis media? A randomised controlled trial. J Paediatr Child Health 2014 May;50(5):362-367. [doi: 10.1111/jpc.12496] [Medline: 24612007]
29. Bramley D, Riddell T, Whittaker R, Corbett T, Lin RB, Wills M, et al. Smoking cessation using mobile phone text messaging is as effective in Maori as non-Maori. N Z Med J 2005 Jun 3;118(1216):U1494. [doi: 10.1186/s13643-017-0520-9] [Medline: 15937529]
30. Sharpe S, Kool B, Whittaker R, Lee AC, Reid P, Civil I, et al. Effect of a text message intervention to reduce hazardous drinking among injured patients discharged from a trauma ward: a randomized controlled trial. NPJ Digit Med 2018;1:13 [FREE Full text] [doi: 10.1038/s41746-018-0019-3] [Medline: 31304298]
31. Yao P, Fu R, Rushing SC, Stephens D, Ash JS, Eden KB. Texting 4 sexual health: improving attitudes, intention, and behavior among American Indian and Alaska native youth. Health Promot Pract 2018 Nov;19(6):833-843. [doi: 10.1177/1524839918761872] [Medline: 29557176]
32. Fletcher R, Hammond C, Faulkner D, Turner N, Shipley L, Read D, et al. Stayin' on track: the feasibility of developing internet and mobile phone-based resources to support young Aboriginal fathers. Aust J Prim Health 2017 Sep;23(4):329-334. [doi: 10.1071/PY16151] [Medline: 28449728]
33. Clark RA, Fredericks B, Buitendyk NJ, Adams MJ, Howie-Esquivel J, Dracup KA, et al. Development and feasibility testing of an education program to improve knowledge and self-care among Aboriginal and Torres Strait Islander patients with heart failure. Rural Remote Health 2015;15(3):3231 [FREE Full text] [Medline: 26390941]
34. Houston A, Laws R, Askew D, Saldanha T, Denney-Wilson E. Exploring the cultural appropriateness and usefulness of a mhealth promotion program for infant feeding in an urban Aboriginal health service: a qualitative study. Aust Indig Health Bull 2017;17(3):1-10 [FREE Full text]
35. Dingwall KM, Puszka S, Sweet M, Nagel T. 'Like drawing into sand': acceptability, feasibility, and appropriateness of a new e-mental health resource for service providers working with Aboriginal and Torres Strait Islander people. Austr Psychol 2015 Jan 12;50(1):60-69. [doi: 10.1111/ap.12100]
36. Kirkham R, MacKay D, Barzi F, Whitbread C, Kirkwood M, Graham S, et al. Improving postpartum screening after diabetes in pregnancy: results of a pilot study in remote Australia. Aust N Z J Obstet Gynaecol 2019 Jun;59(3):430-435. [doi: 10.1111/ajo.12894] [Medline: 30276799]
37. Povey J, Mills PP, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of mental health apps for Aboriginal and Torres Strait Islander Australians: a qualitative study. J Med Internet Res 2016 Mar 11;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
38. Dingwall KM, Puszka S, Sweet M, Mills PP, Nagel T. Evaluation of a culturally adapted training course in indigenous e-mental health. Australas Psychiatry 2015 Dec;23(6):630-635. [doi: 10.1177/1039856215608282] [Medline: 26423096]
39. South Australian Health and Medical Research Institute (SAHMRI). 2019. Alex Brown: Personal Profile URL: https://portal.sahmriresearch.org/en/persons/alex-brown [accessed 2019-05-08] [WebCite Cache ID 78ClzKC3A]
40. Hammond C. The University of Newcastle, Australia. 2017. Deadly Streaming: Identity Learning & Strengths URL: https://www.newcastle.edu.au/__data/assets/pdf_file/0003/468570/Deadly-Streaming-Student-Book-2018-v21.pdf [accessed 2019-05-14] [WebCite Cache ID 78M6HhMP3]
41. Indigenous Allied Health Australia. 2018. IAHA Governance: Our Board URL: https://iaha.com.au/about-us/iaha-board/ [accessed 2019-05-08] [WebCite Cache ID 78Cme1H6C]
42. Institute for Urban Indigenous Health IUIH. 2019. Board and Committees URL: http://www.iuih.org.au/About/Board_and_Committees [accessed 2019-05-08] [WebCite Cache ID 78CqucCEE]
43. Larrakia Nation. 2019. Robbie Mills & The Collective URL: http://larrakia.com/talent-view/robbie-mills-collective/ [accessed 2019-05-08] [WebCite Cache ID 78Cr9UPZ8]
44. Menzies - School of Health Research. 2019. Inspiring Indigenous Researcher URL: https://www.menzies.edu.au/page/Research/Case_studies/Indigenous_health/Inspiring_Indigenous_researcher/ [accessed 2019-05-08] [WebCite Cache ID 78CrQ1iWV]
45. NRWC Muster application - National Rural Womens Coalition. 2016. Lisa Shipley URL: https://www.nrwc.com.au/lisa-shipley [accessed 2019-05-08] [WebCite Cache ID 78CreDt5F]
46. Southcentral Foundation. 2019. Meet the Research Team URL: https://www.southcentralfoundation.com/services/research/meet-the-research-team/ [accessed 2019-05-08] [WebCite Cache ID 78CtGgwkj]
47. Aboriginal and Torres Strait Islander Studies Unit: The University of Queensland. 2019. Professor Bronwyn Fredericks URL: https://atsis.uq.edu.au/profile/772/bronwyn-fredericks [accessed 2019-05-08] [WebCite Cache ID 78CtdbJ6Z]
48. 100 Maori Leaders. 2019. Professor Papaarangi Mary-Jane Reid URL: https://100maorileaders.com/professor-papaarangi-reid [accessed 2019-05-08] [WebCite Cache ID 78CzfabBU]
49. Recruitment Centre for Waitemata District Health Board. 2019. Nurturing and Growing the Maori Health Workforce at Waitemata DHB URL: https://www.wdhbcareers.co.nz/

Nurturing%20and%20Growing%20the%20M%C4%81ori%20Workforce [accessed 2019-05-08] [WebCite Cache ID 78CzY6Vhp]
50. Partnerships for Native Health | Washington State University. 2019. Meghan Jernigan, MPH URL: http://www.p4nh.org/staff/meghan-jernigan/ [accessed 2019-05-08] [WebCite Cache ID 78D0LpVKz]
51. Maar MA, Yeates K, Toth Z, Barron M, Boesch L, Hua-Stewart D, et al. Unpacking the black box: a formative research approach to the development of theory-driven, evidence-based, and culturally safe text messages in mobile health interventions. JMIR Mhealth Uhealth 2016 Jan 22;4(1):e10 [FREE Full text] [doi: 10.2196/mhealth.4994] [Medline: 26800712]
52. Fiordelli M, Diviani N, Schulz PJ. Mapping mhealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
53. Shand F, Mackinnon A, O'Moore K, Ridani R, Reda B, Hoy M, et al. The iBobbly Aboriginal and Torres Strait Islander app project: study protocol for a randomised controlled trial. Trials 2019 Apr 5;20(1):198 [FREE Full text] [doi: 10.1186/s13063-019-3262-2] [Medline: 30953556]
54. Bradford S, Rickwood D. Young people's views on electronic mental health assessment: prefer to type than talk? J Child Fam Stud 2015;24(5):1213-1221 [FREE Full text] [doi: 10.1007/s10826-014-9929-0] [Medline: 25960628]
55. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. J Epidemiol Community Health 2002 Feb;56(2):119-127 [FREE Full text] [doi: 10.1136/jech.56.2.119] [Medline: 11812811]
56. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mhealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
57. Nagel T, Robinson G, Condon J, Trauer T. Approach to treatment of mental illness and substance dependence in remote indigenous communities: results of a mixed methods study. Aust J Rural Health 2009 Aug;17(4):174-182. [doi: 10.1111/j.1440-1584.2009.01060.x] [Medline: 19664081]
58. Menzies - School of Health Research. 2019. Development of the AIMhi Stay Strong App URL: https://www.menzies.edu.au/page/Research/Projects/Mental_Health_and_wellbeing/Development_of_the_Stay_Strong_iPad_App/ [accessed 2019-05-20] [WebCite Cache ID 78WTMFqOV]
59. Dingwall KM, Nagel T, Hughes JT, Kavanagh DJ, Cass A, Howard K, et al. Wellbeing intervention for chronic kidney disease (WICKD): a randomised controlled trial study protocol. BMC Psychol 2019 Jan 8;7(1):2 [FREE Full text] [doi: 10.1186/s40359-018-0264-x] [Medline: 30621791]
60. Sterne JA, Egger M, Moher D. Addressing reporting biases. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. United Kingdom: The Cochrane Collaboration; 2011.
61. Kowal E, Anderson I, Bailie R. Moving beyond good intentions: indigenous participation in aboriginal and Torres Strait Islander health research. Aust N Z J Public Health 2005 Oct;29(5):468-470. [doi: 10.1111/j.1467-842X.2005.tb00228.x] [Medline: 16255450]
62. Lowitja Institute. The Lowitja Institute And The Lowitja Institute Aboriginal and Torres Strait Islander Health CRC Style Guide v14 URL: https://www.lowitja.org.au/page/publishing-policies [accessed 2019-07-19]

Abbreviations

DSI-SS: Depressive Symptom Inventory—Suicidality Subscale
JBI: Joanna Briggs Institute
LoE: level of evidence
MCP: motivational care planning
MMS: multimedia messaging service
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SMS: short messaging service
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 30.05.19; peer-reviewed by T Steeb, J Brütting; comments to author 12.07.19; revised version received 25.07.19; accepted 28.07.19; published 07.10.19.

Please cite as:
Hobson GR, Caffery LJ, Neuhaus M, Langbecker DH
Mobile Health for First Nations Populations: Systematic Review
JMIR Mhealth Uhealth 2019;7(10):e14877
URL: https://mhealth.jmir.org/2019/10/e14877
doi:10.2196/14877
PMID:31593537

Original Paper

Community-Based Chronic Disease Prevention and Management for Aboriginal People in New South Wales, Australia: Mixed Methods Evaluation of the 1 Deadly Step Program

David Peiris[1], MBBS, MPH, PhD; Lachlan Wright[1]; Madeline News[1], BHlthSc, MHC; Katherine Corcoran[1], BSc, MPH

The George Institute for Global Health, UNSW Sydney, Newtown, Australia

Corresponding Author:
David Peiris, MBBS, MPH, PhD
The George Institute for Global Health, UNSW Sydney
1 King St
Newtown, 2042
Australia
Phone: 61 420969204
Email: dpeiris@georgeinstitute.org

Abstract

Background: Chronic diseases account for over 70% of health gaps between Aboriginal people and the rest of the Australian population. The 1 Deadly Step program involves community-based events that use a sporting platform and cultural ambassadors to improve chronic disease prevention and management in New South Wales (NSW).

Objective: This study aimed to evaluate the feasibility and acceptability of a community-based chronic disease screening program for Aboriginal people.

Methods: In 2015, the program was enhanced to include an iPad app for screening assessments, a results portal for nominated care providers, and a reporting portal for program administrators and implemented in 9 NSW community events. A mixed methods evaluation comprising survey data, analytics obtained from iPad and Web portal usage, and key informant interviews was conducted.

Results: Overall, 1046 people were screened between April 2015 and April 2016 (mean age 40.3 years, 640 (61.19%) female, 957 (91.49%) Aboriginal or Torres Strait Islander). High chronic disease rates were observed (231 [22.08%] participants at high cardiovascular disease (CVD) risk, 173 [16.54%] with diabetes, and 181 [17.30%] with albuminuria). A minority at high risk of CVD (99/231 [42.9%]) and with diabetes (73/173 [42.2%]) were meeting guideline-recommended management goals. Overall, 297 participants completed surveys (response rate 37.4%) with 85.1% reporting satisfaction with event organization and information gained and 6.1% experiencing problems with certain screening activities. Furthermore, 21 interviews were conducted. A strong local working group and processes that harnessed community social networks were key to implementation success. Although software enhancements facilitated screening and data management, some technical difficulties (eg, time delays in processing blood test results) impeded smooth processing of information. Only 51.43% of participants had a medical review recorded postevent with wide intersite variability (10.5%-85.6%). Factors associated with successful follow-up included clinic managers with overall program responsibility and availability of medical staff for immediate discussion of results on event day. The program was considered highly resource intensive to implement and support from a central coordinating body and integration with existing operational processes was essential.

Conclusions: 1 Deadly Step offers an effective and acceptable strategy to engage Aboriginal communities in chronic disease screening. High rates of risk factors and management gaps were encountered, including people with no previous knowledge of these issues. Strategies to improve linkage to primary care could enhance the program's impact on reducing chronic disease burden.

(JMIR Mhealth Uhealth 2019;7(10):e14259) doi:10.2196/14259

KEYWORDS

chronic disease; screening; indigenous health; prevention; primary health care

Introduction

Chronic diseases, including cardiovascular disease (CVD), diabetes, chronic kidney disease (CKD), chronic respiratory disease and cancer, account for over 70% of health gaps between Aboriginal people and the rest of the Australian population [1]. Aboriginal and Torres Strait Islander peoples experience around 5 times greater CVD burden than other Australians [1]. Aboriginal people also experience substantial inequities in access to primary health care, and innovative, culturally safe strategies to improve access to high-quality chronic disease care and prevention are needed [2-5]. Studies of CVD risk management in Australian general practice and Aboriginal Community Controlled Health Service (ACCHS) settings demonstrated that 50% of routinely attending adults lacked sufficient recorded information to comprehensively evaluate vascular risk [6]. For those identified at high vascular risk, only around 40% were prescribed guideline-indicated medicines.

Community-based strategies that improve the uptake of best practice recommendations could both substantially reduce the disease burden from chronic diseases and help improve health system efficiencies. The 1 Deadly Step program was developed in partnership with New South Wales (NSW) Health and the Australian Rugby League to address the high prevalence of chronic diseases in NSW Aboriginal communities. The term *deadly* is used by many Aboriginal people to mean *awesome, great, excellent* and the term *1 Deadly Step* makes a link between rugby league and making a step toward good health. First implemented in 2012, it uses a culturally safe, innovative, community-based model in which annual events are held to increase awareness of chronic diseases and to promote prevention, early detection, and evidence-based management of chronic diseases through timely referral and follow-up. At each community event, consenting participants are taken through specific stations to assess chronic disease risk factors. Drawing on the popularity of rugby league in Aboriginal communities, the program uses this sporting platform to encourage local communities to participate. High-profile Aboriginal rugby league players from the local community are engaged as cultural ambassadors and are available on the event day to promote the importance of looking after one's health.

In this paper, we describe the development of an electronic platform to support implementation of the 1 Deadly Step program and outline the findings from a mixed methods evaluation. It draws on the key findings from the full evaluation report prepared for the commissioning agency [7]. The objectives of this study were to (1) describe the demographic and chronic disease risk factor profile of program participants, (2) assess evidence-practice gaps for chronic disease management, and (3) assess program acceptability to both participants and providers and identify implementation barriers and enablers.

Methods

Program Enhancements

An earlier evaluation of the program in 2012 concluded that 1 Deadly Step events represented a successful community screening approach with high acceptability by the participating communities [8]. However, a key recommendation was the need for improvements to screening and data collection processes and support for systematic follow-up care. To help address these issues, an electronic platform was developed comprising 4 components.

iPad Screening App

A screening algorithm was developed as an iPad app. ACCHS staff were engaged to inform the design via a series of workshops and iterative testing of software prototypes. A local Aboriginal artist also provided the software design to improve the visual appeal of the app. The core elements of the system built are shown in Multimedia Appendix 1. It comprises a step-by-step screening process, a printable summary report and a real-time, secure upload of data to a standards-compliant repository.

Administrator Portal

A Web-based portal allows program staff to manage events (Multimedia Appendix 2). This included setting up test events for staff training purposes, registration of staff users responsible for entering data into the iPad on event days, and registration of nominated care providers who would be given access to the provider portal.

Provider Portal

For consenting participants, a clinical summary report is accessible from a password protected, secure data repository. The ACCHS or general practice can view data for the participants who have nominated them as their care provider (Multimedia Appendix 3). Up to 3 care providers can be assigned to each participant (eg, general practitioner [GP], ACCHS manager, and local health district [LHD] staff). The nominated care provider can generate a summary of the screening data and upload this document to the patient's electronic record. The portal has a sort function that prioritizes patients according to their chronic disease risk and follow-up status.

Web-Based Reporting Tool and Site Evaluation Report

Program administrators can access aggregated data reports for each event, which includes the demographic and health profile of participants, numbers of assessments completed, and the nominated care providers. A site evaluation report was also provided by the research team postevent for dissemination to participating health care providers.

Evaluation

The enhanced program was implemented by the NSW Agency for Clinical Innovation in 2015-2016. A program logic model (Multimedia Appendix 4) and evaluation framework using the Reach, Effectiveness, Adoption, Implementation, Maintenance framework was developed by the research evaluation team [9]. This involved detailed discussion with key stakeholders to identify core evaluation objectives and to design strategies and questions that appropriately measured those objectives.

Table 1. Data elements for evaluation of the 1 Deadly Step program.

Data element	Means of collection	Information collected
Screening assessment deidentified data	Secure access to the data repository	Demographic information; clinical information including cardiovascular disease risk, diabetes risk, kidney disease risk, and current treatment being received
Satisfaction surveys for participants	Anonymous paper survey at the end of a screening event	Satisfaction, acceptability, and utility of the program
Reporting website follow-up data	Deidentified data extract at end of project	Follow-up status of all participants screened
Key stakeholder interviews	Semistructured interviews with health service managers, local health district staff, clinical staff, and program staff	Satisfaction, acceptability, and utility of the program

A mixed methods approach was taken to data collection, and 4 main data sources were used to inform the evaluation (Table 1). Deidentified quantitative and identified qualitative data were used concurrently to gain a detailed understanding of the activities, inputs, and outputs of the project as identified in the program logic model.

Risk Factor Analyses

An assessment of the proportion of participants at risk of diabetes, CVD, and CKD was made. CVD risk estimation for those aged 30 years and older was based on the Framingham risk equation and National Vascular Disease Prevention Alliance guidelines [10]. High CVD risk was defined as any of the following: (1) a calculated 5-year CVD risk exceeding 15%, (2) presence of clinically high-risk conditions (including diabetes and age >60 years, diabetes and albuminuria, systolic blood pressure>180 mmHg, diastolic blood pressure >110 mmHg, or total cholesterol >7.5 mmol/L), and (3) a self-reported CVD diagnosis (coronary heart disease, cerebrovascular disease, and peripheral vascular disease). Diabetes risk was based on the Australian type 2 diabetes risk (AUSDRISK) screening tool [11], glycated hemoglobin (HbA_{1c}), and the capillary blood glucose level taken on the event day. CKD risk was defined as any of the following: body mass index (BMI) greater than 30 kg/m^2, current smoker, the presence of CVD, family history of CKD, and the presence of diabetes [12].

Care Practices

The proportion of participants identified with or at high risk of these conditions, who were accessing appropriate management (eg, self-reported use of guideline-recommended medications and attainment of recommended treatment targets) was assessed. Medication use was based on self-reporting. An information pop-up box was available in the iPad app with common medication names to assist in answering these questions.

Participant Satisfaction

At the end of their screening assessment, participants were asked to complete a 2 min survey seeking feedback on the overall event and any problems encountered at each of the screening stations.

Follow-Up

The nominated provider or manager was encouraged to record in the provider portal which participants were followed up and the date of follow-up. Data were extracted from the portal to assess follow-up rates at the end of the program (August 31, 2016).

For the quantitative data, simple frequency analyses were conducted using SAS software version 9.4 (SAS Institute) including assessing for variation by site, gender, and clinical characteristics.

Interviews

Semistructured interviews were conducted with a purposive sample of health service and program staff. A maximum-diversity sampling strategy was taken in which participants were selected on the basis of site, staff role, and involvement in the program. Interviews were generally conducted by telephone, with two evaluation team members taking an insider-outsider approach [13]—one (KC) who was not involved in the program design and the other (LW) an Aboriginal researcher with detailed program knowledge and a long history of engagement with the participating communities. Interviews were digitally recorded, professionally transcribed, and reviewed by a member of the evaluation team to ensure accuracy of the transcription. Thematic analysis was conducted, and themes were aligned with the areas of focus in the logic model. All team members met regularly to develop the coding framework and discuss the significance of the emerging codes and their relevance to the quantitative data that was being concurrently collected. This framework was iteratively revised, and member checking was informally conducted to ensure consistency of interpretation. Future interviews were modified to enable deeper exploration of particular emergent themes.

The evaluation was approved by the Aboriginal Health & Medical Research Council Human Research Ethics Committee. Formal approvals from each of the participating sites were obtained. Informed consent was obtained from all participants who were interviewed.

Results

Participant Characteristics

A total of 1046 people were screened between April 2015 and April 2016 at 9 events in NSW. Table 2 highlights the participant characteristics by site. An average of 116 participants was screened per site, with a larger proportion of females than males screened at all sites (61.2% vs 38.8% overall). The majority of participants (91.5%) identified as Aboriginal or Torres Strait Islander. The mean age of the participants was

40.3 years (range 15-79 years). On the basis of 2016 Indigenous area census data, 5.58% of the population older than 15 years was screened overall (range 3.0%-19.8%).

Data on chronic disease risk factors by gender are summarized in Table 3. For weight-related measures, 50.3% of the sample had a BMI in the obesity range (mean BMI 31.1 kg/m^2) and 65.6% had an elevated waist circumference (>102 cm for men and 88 cm for women and BMI >40 kg/m^2). There were significant gender differences with females recording higher rates of obesity (53.8% vs 44.8%) and elevated waist circumference (76.1% vs 49.0%). For smoking status, 37.2% were current smokers, with a further 6.7% having recently given up smoking in the previous 12 months. Most current smokers had been smoking for more than 10 years (64.1%) and 44.0% smoked more than 10 cigarettes per day. Smoking rates by gender were similar (34.7% males, 38.8% females). Importantly, 20.0% of those under 18 years reported being current smokers.

Data on the CVD risk of participants are summarized in Table 4. Around 1 in 5 (22.08%) of the sample was at high CVD risk, either through having an existing CVD condition or one or more clinically high-risk conditions. There were minimal differences in CVD risk profile by gender. Overall, 16.54% of the participants reported having diabetes (Table 5). An additional 4.30% of the sample had diabetic-range HbA_{1c} levels greater than or equal to 6.5% (48 mmol/mol) without a previous known diagnosis of diabetes. Another 28.87% of participants had potentially elevated glucose levels (random capillary blood glucose levels between 5.5 mmol/L and 11.1 mmol/L).

There were few gender differences in the diabetes risk profile of the sample. In total, 82.98% of participants were at high CKD risk and 17.30% had albuminuria (≥2.5 mg/mmol for males and ≥3.5 mg/mmol for females). There were negligible gender differences in the proportion of people at high risk of CKD overall.

Table 2. Demographic profile of participants screened.

Event site	Date	Number screened	Average age (years)	Female, n (%)	Aboriginal and/or Torres Strait Islander, n (%)	Local Aboriginal and/or Torres Strait Islander community[a], %
1	04/17/2015	132	42.2	85 (64.4)	111 (84.1)	3.6
2	07/06/2015	114	40.6	71 (62.3)	113 (99.1)	3.0
3	10/26/2015	107	34.6	63 (58.9)	90 (84.1)	6.7
4	12/01/2015	118	41.6	71 (60.2)	114 (96.6)	4.5
5	03/06/2016	77	39.0	55 (71.4)	68 (88.3)	3.6
6	03/12/2016	119	38.4	60 (50.4)	114 (95.8)	8.9
7	03/17/2016	123	47.0	73 (59.4)	107 (87.0)	7.0
8	03/23/2016	127	40.0	81 (63.8)	119 (93.7)	11.2
9	04/06/2016	129	37.9	81 (62.8)	119 (93.8)	19.8
Total	—[b]	1046	40.3	61.19	957 (91.49)	5.58

[a]On the basis of 2016 Indigenous area census data for people aged 15 years and older.

[b]Not applicable.

Table 3. Chronic disease risk factors by gender

Risk factors	Female (N=640), n (%)	Male (N=406), n (%)
Current smoker	248 (38.8)	141 (34.7)
Body Mass Index >30 kg/m2	344 (53.8)	182 (44.8)
Elevated waist circumference (>102cm for males, >88cm for females)	487 (76.1)	199 (49.0)
Physical activity less than 2.5 hours/week	139 (21.7)	48 (11.8)
Infrequent fruit intake	185 (28.9)	100 (24.6)
Infrequent vegetable intake	526 (82.2)	337 (83.0)
Blood pressure >140/90 mmHg	200 (31.3)	185 (45.6)
Dyslipidaemia (Total cholesterol > 5.5, HDL <1, LDL >3.5, Triglycerides >2.0)	464 (72.5)	308 (75.9)

Table 4. Cardiovascular and diabetes risk profile for 1046 people.

Cardiovascular disease risk profile	n (%)
Low risk (<10% 5-year risk)	429 (41.01)
Medium risk (10-15% 5-year risk)	17 (1.63)
High risk (>15% 5-year risk)	15 (1.43)
Clinically high-risk condition present	79 (7.55)
Established cardiovascular disease	137 (13.10)
<30-year olds	312 (29.83)
Missing data	57 (5.45)

[a]Australian type 2 diabetes risk screening assessment.

Table 5. Diabetes risk profile for 1046 people.

Diabetes risk profile	n (%)
Low risk (AUSDRISK[a] ≤6)	31 (2.96)
Medium risk (AUSDRISK 6-11)	145 (13.86)
High risk (AUSDRISK ≥12)	350 (33.46)
Impaired glycemia	302 (28.87)
Possible new diabetes diagnosis	45 (4.30)
Established diabetes	173 (16.54)

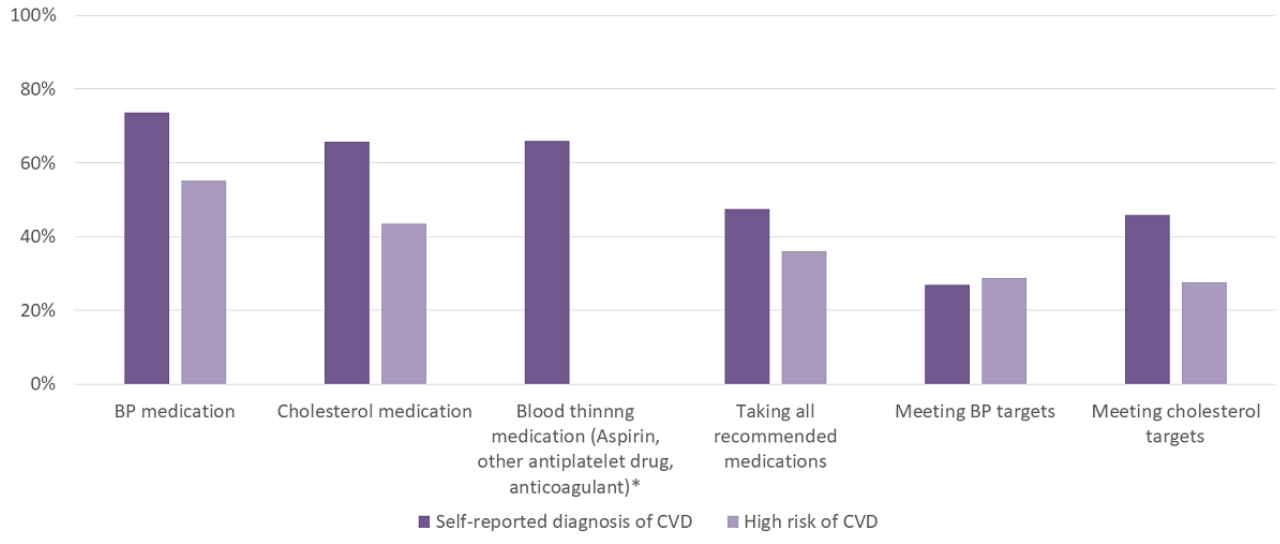
[a]Australian type 2 diabetes risk screening assessment.

Figure 1 illustrates the proportion of patients with CVD and at high risk of CVD meeting various care practice parameters. Overall, only 42.9% of people with or at high risk of CVD were taking guideline-recommended treatments. There were no significant gender differences in those who reported currently taking these medicines.

For those with a known diagnosis of diabetes (n=173), the majority (80.4%) reported taking an oral glucose-lowering medication and 34.7% reported taking insulin. Overall, 42.2% were attaining a target HbA_{1c} of 7% or less (53 mmol/mol) and 61.9% were attaining a target HbA_{1c} of 8% or less (64 mmol/mol). A higher proportion of women met the HbA_{1c} targets than men (46.3% vs 35.4%, respectively).

As of August 2016, 538 of the 1046 participants screened had been recorded as having been followed up (51.43%). There was wide variability in the recording of follow-up rates (10.5%-85.6%). Participants with or identified to be at high risk of diabetes, CVD, or CKD had slightly higher follow-up rates than the total population at each site and overall.

Figure 1. Management for people with or at high risk of CVD (n=231). BP: blood pressure; CVD: cardiovascular disease.



Satisfaction

A total of 297 participants completed satisfaction surveys at 7 events (response rate 37.4%). The overall impressions of the program were positive, with the vast majority of participants satisfied with event organization and the information gained from the event day (Figure 2). Similarly, the majority of responders encountered few problems when asked about specific screening stations, although around 6% of participants did report problems with the urine and blood testing stations. Free-text entries were also analyzed and were concordant with these findings.

Figure 2. Participant survey of overall program impression (n=297). 1 DS: 1 Deadly Step.

Interviews

A total of 21 interviews were conducted (Table 6). Interview themes have been organized to align with the three key stages of the program (pre-event, event, and postevent).

Pre-Event

Importance of the Working Group

An important initial stage in organizing an event involved the coordinating agency engaging local stakeholders to determine interest and capacity. A stakeholder working group was responsible for pre-event planning, estimation of staffing and equipment requirements, engagement with Country Rugby League to identify ambassadors, running the event, and determining the follow-up processes for participants. The working group was viewed favorably by interviewees. One ACCHS Chief Executive Officer (CEO) commented that "getting all the main players (together)... worked really well." Interviewees expressed positive sentiments about the coordinating agency's administration and management of the working group, with 1 ACCHS project officer stating she "couldn't fault this part of the support." The model of collaboration was seen as a blueprint for a broader range of health promotion activities:

> *...the staff just came in droves and…having those meetings was just so great. That's the way we should do all of our health promotion.* [ACCHS practice manager]

Table 6. Interviews by professional category.

Site	Chief executive officer	Clinic manager	General practitioner	Nurse care coordinator	Aboriginal project or liaison officer	Local Hospital District staff	Program staff	Other	Total
1	—[a]	1	—	1	1	—	—	—	3
2	1	—	—	—	1	1	—	—	3
3	—	1	—	1	—	2	—	—	4
4	1	—	—	1	—	2	—	1	5
5	—	—	—	—	—	—	—	—	0
6	—	—	—	1	1	—	—	—	2
7	—	—	1	—	—	—	—	—	1
8	—	—	—	—	—	—	—	—	0
9	—	—	—	1	—	—	—	—	1
Other	—	—	—		—	—	2	—	2
Total	2	2	1	5	3	5	2	1	21

[a]Cells with dashes indicate that no professionals were interviewed in that category.

Staff Training

Interviewees considered a lead time of around 3 months was required to adequately plan for an event. In general, pre-event staff training was considered sufficient and was usually conducted as a one-off, half-to-whole-day activity supplemented by training on the event day itself. Staff participation at training events required support from managers and advanced planning to free up staff time for attendance. One interviewee felt it was "up to management pushing that need for people to attend." At some sites, however, the inability to do pre-event training did not appear to be a major barrier:

> *no training had been done with the app at all [due to technical difficulties on the training day]... and so we winged it on the day...it took probably about a minute with each person and everyone had the hang of it. So that just proves how simple the app is to use I guess.* [LHD coordinator]

One manager commented that the training could be more structured and that less experienced staff might benefit from small group learning. Several participants also suggested dedicating more training in the use of the point-of-care machinery.

Event Day

Implementing a Clinical Program in a Community Setting

Interviewees were generally positive about event day, viewing it as an opportunity for family and friends to come together. Holding events outdoors rather than inside a clinic facility was seen as particularly important:

> *it's a good get together...you can't put money on the value of community getting together... having a yarn and catching up with each other.* [ACCHS registered nurse]

It was also perceived to be a fun way to learn more about risk factors for chronic disease. One hospital district nurse felt that it went beyond the usual "deficit" model of Aboriginal health and served to empower individuals:

> *Aboriginal people must get... bored of the statistics thrown at them about chronic disease...and their lifestyle management is causing all of these problems. Whereas I don't think that's what 1 Deadly Step did. I think it was a really positive way of getting the message across that you can do something about this and we're here to help you.* [LHD Clinical Nurse Consultant]

It also gave staff an opportunity to strengthen relationships with staff from other organizations. These events also have potential to boost ACCHS staff morale and make their work more visible to board members. Some events were held concurrently with a longstanding national cultural event (National Aborigines and Islanders Day Observance Committee, NAIDOC) and this was an effective strategy to *demedicalize* the program:

> *NAIDOC draws everybody…They don't really see it as coming here to get screened. They see it as—if I do all these steps I get that cool jersey and I get to have a feed and I get to have a day out with my family.* [ACCHS practice manager]

However, running an event alongside another community event also increased operational complexity. Aboriginal staff, in particular, have community and family responsibilities at these events in addition to their work responsibilities. This appeared to make management and oversight of the day more difficult. Despite many participants highlighting the importance of holding outdoor events, inclement weather also poses additional challenges such as exposed electrical cords, marquees becoming unstable in the wind, and an increased potential for biological samples to be incorrectly processed, misplaced, or tipped over.

Work Flow Considerations

Most people considered the clinical information collected to be important; however, this needed to be balanced against managing the workflow associated with large-scale screening. Consequently, there were mixed views concerning the optimal amount of information that should be collected. Some interviewees considered that all of the data were important as they could support other service activities such as completion of government-rebated Aboriginal health assessments. One ACCHS manager also considered this information of particular importance for improving the quality of their key performance indicator data that is provided to funding bodies. Other interviewees questioned the relative merits of conducting point-of-care testing for cholesterol, diabetes, and kidney disease for all participants (as discussed further).

Role of Country Rugby League and Marketing Activities

The use of Country Rugby League ambassadors was considered as a useful community engagement strategy, making the event more fun for children and freeing up parent or carer time for screening. Some sites used the ambassadors to leverage additional marketing opportunities through free local media advertising or via a Facebook page. Although viewed favorably, some interviewees commented that the ambassadors need to be more committed to supporting the program:

> *there's a role for ambassadors, but...if...the ambassadors can't speak passionately about it, people can see straight through that...Often it's sold as the rugby league player's going to be there and then they're not...it's like a con job.* [ACCHS CEO]

1 Deadly Step shirts were also critically important incentives to enhance event attendance:

> *in all the years I've been working in Aboriginal health, shirts are a really big incentive, people...love to wear them and they love to promote them.* [ACCHS clinic manager]

Encouraging the staff to wear the event shirts in the weeks leading to the event was a successful marketing strategy, as patients were asking "How do I get one of the shirts?" Some participants were disappointed that the shirt design remained unchanged from the previous year, further highlighting the importance of refreshing the designs regularly.

Technical Challenges

There was general consensus that the iPad app was easy to use and required minimal training. The main issue raised by some interviewees was related to the challenges of entering results at the blood test station. Additional problems included an inability to enter an error code when a patient's result was outside the range of the machine. Some interviewees suggested taking a modular approach to screening events where event organizers could adapt the program to their specific requirements. This included provision of a *light* version of the app that did not include all of the mandatory screening stations.

Several interviewees raised the issue of bottlenecks associated with the blood and urine testing stations. The point-of-care machines take several minutes to process a result, and problems particularly arose when samples had to be rerun because of errors:

> *...it's like the brake lights on the highway...Once you start having to...rerun samples and potentially getting the same error…it slows down the flow of people going through.* [ACCHS CEO]

Some interviewees reported problems in printing patient reports, which was related to the use of older model iPads. Other interviewees commented on problems related to insufficient network capacity and problems with printers. These delays resulted in some participants leaving before receiving their report or having a discussion with GPs and nurses.

Postevent

Challenges With Follow-Up

Most interviewees considered follow-up activities to be resource intensive. At 1 high-performing site with follow-up rates over 80%, the ACCHS clinic manager reflected that it was around a two-month process to implement adequately:

> *It was six to eight weeks, and there are still people with low level risk...and we're still capturing them...But all the ones that we listed as priority, have been followed up.* [ACCHS clinic manager]

One ACCHS CEO felt that these requirements should be made clearer in the working group when preparing for an event:

> *...there's a lot of work in the event, but potentially there's a heap of work after the event and you need to think through what your follow up strategies are going to be...* [ACCHS CEO]

Much of the follow-up processes were implemented by managers, administrators, and Aboriginal Health Workers. At sites with staff shortages, these processes were particularly difficult to operationalize. The major challenge was related to participants who nominated a care provider other than the local ACCHS for follow-up. For non-ACCHS participants, it was originally envisaged that the nominated GPs would be registered into the system and notified of the patients who had requested follow-up through them. However, in practice, this process was difficult to implement, and consequently, the coordinating agency modified the process such that a hospital district staff member was given responsibility for follow-up of non-ACCHS participants.

Coordinating follow-up processes between different sectors was seen as a valuable outcome from LHD participation in the program. However, some staff suggested this needed to translate into more tangible benefits to justify their participation. For example, 1 LHD staff member would have liked the follow-up process to go one step further and allow for uploads of patient reports into the hospital record system:

> *...I suppose the question is what was in it for us?...quite a lot of staff were involved–paid for by the LHD...if we didn't get any access to the results and be able to have some input into those patient's care then why would we be involved in the first place?* [LHD care coordinator]

One solution to improve follow-up care at some sites was availability of GPs and/or senior nurses on the event day itself for immediate discussion with participants. These sites ensured sufficient privacy for the participant, and it reduced the managerial and administrative staff workload postevent.

Implementing Population Management Processes

The principal process for follow-up was through the use of the provider-reporting portal. Although not all providers consistently used the portal, those that had used it were generally positive and appreciated its simplicity. In addition to the patient-specific reports, the overall event report summary that was provided to key stakeholders was also generally viewed positively:

> *I thought it [the event report Summary] was really good...it gives you exactly how many people were screened, how many people nominated the stakeholders as their provider...It highlighted for us as a health provider that you can target programs around some of that data* [[ACCHS project officer]

Although this report was also reviewed by the working groups, it was unclear to what extent the information was used to inform population health activities. One ACCHS CEO felt that they could benefit from strategic advice on implementation of chronic disease management programs across their community in light of the findings:

> *it would be useful for somebody to work with the services around what they might do with that information...these are your risk factors in the community...and do some projections around...where it could head.* [ACCHS CEO]

Several interviewees also commented on the need for greater integration of the data into routine service provision. This included having the ability to upload participant reports directly into the electronic health records and to generate referrals for specific services such as smoking cessation. The current systems allowed only for uploading of static documents into the patient file. There was a strong interest in being able to upload results into the coded fields of the patient record that could then be used to autopopulate items required for key performance indicator reports and for Medicare-rebated health assessments.

Sustainability

Although stakeholders were generally enthusiastic about involvement in 1 Deadly Step, the trade-off between investing in this program and other activities was raised, particularly by non-ACCHS interviewees:

> *...they're really pushing hard for activity-based funding...we certainly had discussions early on about whether it would be possible (to justify)...the time away from people's usual...work activities.* [LHD Clinical Nurse Consultant]

The funding provided to sites was generally considered insufficient to cover actual costs, and it appears that a considerable amount of in-kind support (eg, local companies providing free generators) was harnessed to support implementation. Some interviewees suggested that multiple stakeholders should pool resources from existing budgets to rationalize costs:

> *...we weren't trying to delegate to each other which can cause controversy with who's bossing who...We knew exactly what we had to do and, with that, we found we were more inclined to give more back into it.* [LHD coordinator]

An additional factor that may influence sustainability was the need to establish a system whereby shared learnings from sites could be made available to other sites. One interviewee, for instance, raised the idea that an experienced person from a site that had held a 1 Deadly Step event earlier might mentor those responsible for conducting an event elsewhere.

Discussion

Principal Findings

In this paper, we examined multiple data sources to evaluate a community-based chronic disease screening and management program for Aboriginal communities in 1 Australian state. There are three main findings from the evaluation: (1) the clinical profile of participants suggested a high burden of chronic diseases and their risk factors, (2) the program had high satisfaction and acceptability rates with several implementation barriers and enablers identified, and (3) factors that might influence the maintenance and sustainability of the program were observed. These findings have important implications for future iterations of the program.

It is difficult to determine representativeness of the communities in which events were held as people often travel large distances to attend. Therefore, we used a large geographic boundary to determine the draining population and estimated that 6% of the population were screened at these events. When compared with the 2012-2013 Australian Aboriginal and Torres Strait Islander Health Survey (AATSIHS) several observations can be made of the 1 Deadly Step participant profile [14,15]. In terms of lifestyle behaviors, current smoking rates and elevated waist circumference were similar in 1 Deadly Step compared with the AATSIHS. Combined overweight and obesity rates (74% vs 66%), elevated blood pressure (37% vs 20%), dyslipidemia (74% vs 51%), CKD rates (21% vs 17%), and diabetes (21% vs 11%) were all higher in 1 Deadly Step compared with the AATSIHS. Despite being a predominantly nonremote sample, these elevated rates are closer to those observed in people from remote areas in the AATSIHS. The prevalence of self-reported CVD in 1 Deadly Step was similar to those reporting heart disease in the AATSIHS (13% vs 12%). The rates of taking recommended treatments for those with or at high risk of CVD were low (47% for CVD and 36% for high CVD risk) and consistent with previous studies [6,16,17].

There are two important implications from the risk factor and care management information. First, 1 Deadly Step is a useful strategy for identifying people at high risk of chronic disease. To enhance the reach of the program, repeated events are likely to be needed to increase the proportion of eligible community members participating. The risk factor prevalence rates of 1 Deadly Step participants is considerably higher and occurs at younger ages than for non-Indigenous people, and in several areas, these rates are higher than reported in representative surveys of Aboriginal and Torres Strait Islander people. There are clearly substantial opportunities to use the program to address access barriers for this group. Second, substantial gaps in optimal care for those with or at high risk of chronic diseases were observed, and consequently, there are also major opportunities for providing higher quality care for these groups through better linkages with their primary care providers.

Several factors influenced program implementation. Participants were generally highly satisfied with the program, and staff involved in its implementation were positive about their involvement. The ability of 1 Deadly Step to draw on existing *community capital* is perhaps the strongest asset of the program. Implementation of traditional clinical processes into such a setting requires considerable planning and harnessing of resources, and the ability to effectively engage communities in this process is noteworthy. The use of existing community events such as NAIDOC, judicious marketing through local and social media, involvement of country rugby league ambassadors, and coordination of activities via a local working group were highly effective strategies to support program participation. Staff from all health service sectors and levels within their organizations demonstrated immense enthusiasm for conducting activities that could address the high–chronic disease burden experienced by Aboriginal communities. Staff welcomed the opportunity to work collaboratively and showcase their efforts to the community. The coordinating agency played a critical role in supporting implementation of the program. Given the resource constraints under which these staff were working, this level of support for the program was an essential enabler to its successful implementation. The ability to be locally responsive while at the same time providing a *macrolevel* view of the program is an important success factor.

A number of areas were also identified where the program was hindered by specific challenges. The most substantial issue was related to the follow-up of participants, particularly those who nominated another service provider other than the local ACCHS. Follow-up rates were highly variable across sites and suggest that there are particular local service issues at play. This has been documented in previous research conducted in the Northern Territory and Queensland [18]. Services with high follow-up rates appear to have committed substantial internal resources

to supporting follow-up. Some sites were able to effectively integrate 1 Deadly Step into existing operational processes and used it as an opportunity to strengthen their reporting capacity to funders. This highlights the challenge of intersector collaboration between local hospital districts, private GPs, ACCHSs, and other agencies, given that there are different jurisdictional responsibilities, information systems, and care processes across these stakeholder groups.

There were also a range of technical hurdles that needed to be overcome. These appeared related to the use of point-of-care machines, network connectivity, and entering data on the iPad software itself. Aside from fixing technical bugs, a major implication is determining the most feasible amount of clinical information that can be collected. Some software modifications could be considered to reduce the amount of information collected at an event. A modular approach could be taken to screening where only certain high-risk subgroups need to have blood and urine testing. A "low-information" algorithm for risk prediction could also be used that was less reliant on complete risk factor information, and this could considerably shorten screening time and reduce bottlenecks. Such algorithms have been developed and validated in overseas populations but would likely need some adaptation and validation work before being used in this setting [19]. Balanced against this, however, is that high–risk factor prevalence rates observed in this program were apparent for a large proportion of the population screened. This provides a strong justification for comprehensive screening. Thus, there is an inherent tension between collecting sufficient information to make the clinical assessment meaningful and overburdening services and participants on the event day.

There are several opportunities that could be derived from 1 Deadly Step. The most important is to integrate 1 Deadly Step data into service processes, thus making it a part of routine business planning and operations [20]. Integration with electronic health record systems would allow providers easy access to screening assessment data in much the same way as specialist and pathology reports are currently viewed, actioned by staff, and extracted for auditing and recall processes [21]. This would also enhance reporting requirements to funders; increase capacity to meet key performance indicator requirements; drive referral to other services; and generate business revenue opportunities, such as meeting the requirements for government-rebated preventive health assessments [22]. Another important opportunity to build on the strengths of the 1 Deadly Step program is to foster collaboration and learning between communities that have participated in events. Currently, these events are run in a stand-alone manner; however, it would be worth considering the establishment of a *learning collaborative* in which nominated representatives could share resources, determine optimal operational approaches, benchmark their performance, and develop data-driven strategies to address follow-up challenges [23]. As consumer-controlled electronic health records become more prominent, integration of screening assessments into these systems could also enhance engagement and promote better exchange of information between care providers [24].

Limitations

This evaluation comes with some important limitations. The sample attending 1 Deadly Step events was a relatively small convenience sample, and therefore, it may not be representative of the local communities participating. The 9 events were from a variety of locations across rural and urban settings in 1 Australian state; therefore, they may not be generalizable to other parts of Australia. The satisfaction survey response rate was low and could be prone to respondent bias. The assessment of the use of guideline-recommended medications was based on self-reported data that is subject to recall bias [25]. The information on follow-up visits was based on reporting activity in the Web portal. Although all health services were regularly encouraged to keep this information up to date, there may be some attendees who received follow-up care, but this was not recorded in the portal.

Taking into consideration some of the existing strengths of the current program, the major challenges faced by 1 Deadly Step relate to its ability to sustain current levels of activity and scale-up activity across NSW. Considerable effort was expended to run these 9 events, and there is a risk that the goodwill associated with this may diminish, particularly as competing demands may displace 1 Deadly Step for other higher-priority areas. Therefore, it is important that the value proposition to staff and stakeholder organizations be high [26]. Real program costs need to be assessed to allow stakeholder organizations to gain a better understanding of the resource requirements to implement 1 Deadly Step. In parallel, it is also important to assess any downstream impact the program may have in terms of savings to the health system made through earlier intervention for chronic disease risk factors and management. An economic evaluation that links 1 Deadly Step participants to other routinely available datasets may aid in addressing this question.

Conclusions

1 Deadly Step was implemented in 9 communities in 2015-2016 and assessed the chronic disease risks for over 1000 Aboriginal people residing in these communities. The clinical data strongly support the justification for such a program given the high levels of risk factors encountered, often including people who would otherwise have had no knowledge of these issues before the events. Overall, the event implementation was highly successful and demonstrated high satisfaction by participants and staff alike. However, several challenges were highlighted, particularly in relation to resource constraints and follow-up processes. Several opportunities were identified to address these issues, and these are likely to play a critical role in influencing program sustainability. It is important to note that despite the successful implementation of the program, its effects in improving health outcomes remain unknown and a more detailed impact evaluation with long-term follow-up is needed to assess downstream system and health benefits.

Acknowledgments

The authors would like to thank all of the communities and organizations participating in 1 Deadly Step, particularly those individuals that participated in the interviews. The authors also thank Kiel Hennesey and Scott Trindal from the NSW Agency for Clinical Innovation for their support with the evaluation.

Authors' Contributions

DP and LW designed the evaluation. LW, MN, and KC were responsible for data collection. All authors participated in the analysis. DP wrote the first draft of the manuscript, and all authors provided critical input to the manuscript and approved its final submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Samples of the screening application.
[PNG File , 126 KB - 1deadlystepSup1.png]

Multimedia Appendix 2
Administrator portal.
[PNG File , 1929 KB - 1deadlystepSup2.png]

Multimedia Appendix 3
Provider portal.
[PNG File , 2251 KB - 1deadlystepSup3.png]

Multimedia Appendix 4
Logic model.
[PNG File , 4986 KB - 1deadlystepSup4.png]

References

1. Vos T, Barker B, Stanley L, Lopez AD. The Burden of Disease and Injury in Aboriginal and Torres Strait Islander Peoples 2003. Brisbane, Australia: The University of Queensland; 2007.
2. Peiris D, Brown A, Howard M, Rickards BA, Tonkin A, Ring I, et al. Building better systems of care for Aboriginal and Torres Strait Islander people: findings from the Kanyini health systems assessment. BMC Health Serv Res 2012 Oct 28;12:369 [FREE Full text] [doi: 10.1186/1472-6963-12-369] [Medline: 23102409]
3. Kelaher M, Dunt D, Thomas D, Anderson I. Comparison of the uptake of health assessment items for Aboriginal and Torres Strait Islander people and other Australians: implications for policy. Aust New Zealand Health Policy 2005 Sep 9;2:21 [FREE Full text] [doi: 10.1186/1743-8462-2-21] [Medline: 16150154]
4. Si D, Bailie R, Cunningham J, Robinson G, Dowden M, Stewart A, et al. Describing and analysing primary health care system support for chronic illness care in indigenous communities in Australia's northern territory - use of the chronic care model. BMC Health Serv Res 2008 May 28;8:112 [FREE Full text] [doi: 10.1186/1472-6963-8-112] [Medline: 18505591]
5. Cass A, Lowell A, Christie M, Snelling PL, Flack M, Marrnganyin B, et al. Sharing the true stories: improving communication between Aboriginal patients and healthcare workers. Med J Aust 2002 May 20;176(10):466-470. [doi: 10.5694/j.1326-5377.2002.tb04517.x] [Medline: 12065009]
6. Peiris D, Patel A, Cass A, Howard M, Tchan M, Brady J, et al. Cardiovascular disease risk management for Aboriginal and Torres Strait Islander peoples in primary health care settings: findings from the Kanyini Audit. Med J Aust 2009 Sep 21;191(6):304-309. [Medline: 19769551]
7. Peiris D, Wright L, Corcoran K, News M, Turnbull F. Excellence and Innovation in Healthcare | EIH - NSW Government. 2016. 1 Deadly Step: Process Evaluation of a Chronic Disease Screening Program in NSW Aboriginal Communities URL: http://www.eih.health.nsw.gov.au/__data/assets/pdf_file/0006/386736/1-Deadly-Step-Process-evaluation-of-a-chronic-disease-screening-program-in-NSW-Aboriginal-communities.pdf [accessed 2019-09-02] [WebCite Cache ID 77zfrfzDL]
8. Sanson-Fisher R, Paul C, Bryant J, Noble N. Excellence and Innovation in Healthcare | EIH - NSW Government. 2012. One Deadly Step Evaluation Report: All Sites URL: http://www.eih.health.nsw.gov.au/__data/assets/pdf_file/0005/386798/One-Deadly-Step-2012-Evaluation-Report.pdf [accessed 2019-09-03] [WebCite Cache ID 77zg35zRy]
9. Gaglio B, Shoup JA, Glasgow RE. The RE-AIM framework: a systematic review of use over time. Am J Public Health 2013 Jun;103(6):e38-e46. [doi: 10.2105/AJPH.2013.301299] [Medline: 23597377]

10. Australian Absolute Cardiovascular Disease Risk Calculator. 2012. Guidelines for the Management of Absolute Cardiovascular Disease Risk URL: http://www.cvdcheck.org.au/pdf/Absolute_CVD_Risk_Full_Guidelines.pdf [accessed 2019-09-03] [WebCite Cache ID 77zfi3c7q]
11. Chen L, Magliano DJ, Balkau B, Colagiuri S, Zimmet PZ, Tonkin AM, et al. AUSDRISK: an Australian type 2 diabetes risk assessment tool based on demographic, lifestyle and simple anthropometric measures. Med J Aust 2010 Feb 15;192(4):197-202. [doi: 10.5694/j.1326-5377.2010.tb03507.x] [Medline: 20170456]
12. Kidney Health Australia. 2015. Chronic Kidney Disease (CKD) Management in General Practice: Guidance and Clinical Tips to Help Identify, Manage and Refer Patients With CKD in Your Practice URL: https://kidney.org.au/cms_uploads/docs/ckdm-in-gp-handbook.pdf [accessed 2019-04-29] [WebCite Cache ID 77zZYxx7f]
13. Dwyer SC, Buckle JL. The space between: on being an insider-outsider in qualitative research. Int J Qual 2009 Mar;8(1):54-63. [doi: 10.1177/160940690900800105]
14. Australian Bureau of Statistics. 2014. Australian Aboriginal and Torres Strait Islander Health Survey: Biomedical Results, 2012-13 URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4727.0.55.003~2012-13~Main%20Features~Chronic%20Kidney%20Disease~113 [accessed 2019-09-03] [WebCite Cache ID 77zQPtsmT]
15. Australian Bureau of Statistics. 2014. Australian Aboriginal and Torres Strait Islander Health Survey: Updated Results, 2012–13 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/4727.0.55.006 [accessed 2019-09-03] [WebCite Cache ID 77zZQctPv]
16. Peiris D, Usherwood T, Panaretto K, Harris M, Hunt J, Redfern J, et al. Effect of a computer-guided, quality improvement program for cardiovascular disease risk management in primary health care: the treatment of cardiovascular risk using electronic decision support cluster-randomized trial. Circ Cardiovasc Qual Outcomes 2015 Jan;8(1):87-95. [doi: 10.1161/CIRCOUTCOMES.114.001235] [Medline: 25587090]
17. Patel A, Cass A, Peiris D, Usherwood T, Brown A, Jan S, Kanyini Guidelines Adherence with the Polypill (Kanyini GAP) Collaboration. A pragmatic randomized trial of a polypill-based strategy to improve use of indicated preventive treatments in people at high cardiovascular disease risk. Eur J Prev Cardiol 2015 Jul;22(7):920-930. [doi: 10.1177/2047487314530382] [Medline: 24676715]
18. Matthews V, Burgess CP, Connors C, Moore E, Peiris D, Scrimgeour D, et al. Integrated clinical decision support systems promote absolute cardiovascular risk assessment: an important primary prevention measure in Aboriginal and Torres Strait Islander primary health care. Front Public Health 2017;5:233 [FREE Full text] [doi: 10.3389/fpubh.2017.00233] [Medline: 28929097]
19. World Health Organization. International Society of Hypertension. 2007. WHO/ISH Risk Prediction Charts For 14 WHO Epidemiological Sub-Regions URL: http://ish-world.com/downloads/activities/colour_charts_24_Aug_07.pdf [accessed 2019-09-04]
20. Wakerman J, Humphreys JS, Wells R, Kuipers P, Entwistle P, Jones J. Primary health care delivery models in rural and remote Australia: a systematic review. BMC Health Serv Res 2008 Dec 29;8:276 [FREE Full text] [doi: 10.1186/1472-6963-8-276] [Medline: 19114003]
21. Matthews V, Burgess CP, Connors C, Moore E, Peiris D, Scrimgeour D, et al. Integrated clinical decision support systems promote absolute cardiovascular risk assessment: an important primary prevention measure in Aboriginal and Torres Strait Islander primary health care. Front Public Health 2017;5:233 [FREE Full text] [doi: 10.3389/fpubh.2017.00233] [Medline: 28929097]
22. Spurling GK, Askew DA, Schluter PJ, Hayman NE. Implementing computerised Aboriginal and Torres Strait Islander health checks in primary care for clinical care and research: a process evaluation. BMC Med Inform Decis Mak 2013 Sep 21;13:108 [FREE Full text] [doi: 10.1186/1472-6947-13-108] [Medline: 24053425]
23. O'Donoghue L, Percival N, Laycock A, McCalman J, Tsey K, Armit C, et al. Evaluating Aboriginal and Torres Strait Islander health promotion activities using audit and feedback. Aust J Prim Health 2014;20(4):339-344. [doi: 10.1071/PY14048] [Medline: 25342123]
24. Neubeck L, Coorey G, Peiris D, Mulley J, Heeley E, Hersch F, et al. Development of an integrated e-health tool for people with, or at high risk of, cardiovascular disease: the consumer navigation of electronic cardiovascular tools (CONNECT) web application. Int J Med Inform 2016 Dec;96:24-37. [doi: 10.1016/j.ijmedinf.2016.01.009] [Medline: 26847070]
25. Stewart J, Sanson-Fisher R, Eades S. Aboriginal and Torres Strait Islander health: accuracy of patient self-report of screening for diabetes, high cholesterol and cervical cancer. Aust N Z J Public Health 2016 Apr;40(Suppl 1):S3-S6. [doi: 10.1111/1753-6405.12309] [Medline: 25716071]
26. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]

Abbreviations

AATSIHS: Australian Aboriginal and Torres Strait Islander Health Survey

ACCHS: Aboriginal Community Controlled Health Service
BMI: body mass index
CEO: chief executive officer
CKD: chronic kidney disease
CVD: cardiovascular disease
GP: general practitioner
HbA_{1c}: glycated hemoglobin
LHD: local hospital district
NAIDOC: National Aborigines and Islanders Day Observance Committee
NSW: New South Wales

Edited by G Eysenbach; submitted 20.05.19; peer-reviewed by J Lynch, S Hinder; comments to author 16.07.19; revised version received 23.07.19; accepted 18.08.19; published 21.10.19.

Please cite as:
Peiris D, Wright L, News M, Corcoran K
Community-Based Chronic Disease Prevention and Management for Aboriginal People in New South Wales, Australia: Mixed Methods Evaluation of the 1 Deadly Step Program
JMIR Mhealth Uhealth 2019;7(10):e14259
URL: http://mhealth.jmir.org/2019/10/e14259/
doi:10.2196/14259
PMID:31638591

Original Paper

Co-designing a Mobile Gamified Attention Bias Modification Intervention for Substance Use Disorders: Participatory Research Study

Melvyn Zhang[1], MBBS, MRCPsych; Sandor Heng[1], DPsych; Guo Song[1], PhD; Daniel SS Fung[2], MBBS, MMED; Helen E Smith[3], DM, FFPHM

[1]National Addictions Management Service, Institute of Mental Health, Singapore, Singapore

[2]Department of Developmental Psychiatry, Institute of Mental Health, Singapore, Singapore

[3]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Melvyn Zhang, MBBS, MRCPsych
National Addictions Management Service
Institute of Mental Health
10 Buangkok Green Medical Park
Singapore, 539747
Singapore
Phone: 65 63892000
Email: melvynzhangweibin@gmail.com

Abstract

Background: Advances in experimental psychology have highlighted the need to modify underlying automatic cognitive biases, such as attentional biases. The effectiveness of bias modification has been well studied for substance use disorders. With recent advances in technology, it is now possible to work outside the laboratory with Web-based and mobile-based attention bias interventions. Gamification technologies might also help diminish the repetitiveness of the task and increase the intrinsic motivation to train. The inconsistent findings of the impact of gaming on the effectiveness of mobile interventions call for further work to better understand the needs of patients (users) and health care professionals.

Objective: The aim of this study was to involve patients, together with health care professionals, in the design of a gamified mobile attention bias modification intervention for substance use disorders.

Methods: The participatory design research method adopted is that of a user-oriented design approach in the form of a future workshop. In the first phase of the workshop, participants shared their critique of an attention bias modification intervention. In the second phase of the workshop, participants were asked to brainstorm features. Participants were also shown gamification approaches and asked to consider if gaming elements could enhance the existing app. In the last phase, participants were asked to sketch a new prototype.

Results: Three co-design workshops were conducted with health care professionals, inpatients, and outpatients. There were 20 participants, consisting of 10 health care professionals and 10 patients. When asked to identify the limitations in the existing app, common issues identified were those of the design, visual probe task, and the included images. Outpatients were also concerned with the safety of administration of the intervention. In the brainstorming sessions, health care professionals made recommendations as to how the stimulus, the mechanism of responding, and the presentation of the scores could be enhanced. Inpatient participants recommended the addition of functionalities, such as information on the harms associated with the substance use, and for there to be enhancements in the design, images, and task. Outpatient participants perceived a need to improve the images and presentation of the results and recommended the inclusion of gaming features. There were differences in opinion on the inclusion of gaming features, as only health care professionals endorsed their inclusion. In the last phase of the workshop, participants were tasked with the conceptualization of prototypes, and the commonality in the design was for a gradual shortening of the interval for stimulus/image presentation.

Conclusions: The results from this research will guide the development of an app that meets the specific needs of patients and is still based on a pre-existing validated task paradigm.

(JMIR Mhealth Uhealth 2019;7(10):e15871) doi:10.2196/15871

KEYWORDS

attention bias; cognitive bias; gamification; participatory design research; psychiatry; apps; cognitive bias; mobile intervention

Introduction

Illicit substances like opioids, cannabis, and stimulants are highly abused worldwide [1]. Unfortunately, there are limited pharmacological approaches to help the affected individuals achieve and maintain abstinence, which is crucial because of the potentially severe complications such as medical comorbidities and death [2,3]. Therefore, the mainstay of management is psychological therapies. Although cognitive behavioral therapy is used often, its effectiveness is variable and not always sustained; some studies reported that 40%-50% of individuals relapsed in the first year and 70% relapsed within 3 years [4]. It appears that conventional psychotherapies do not address all the etiological factors; hence, individuals relapse back to addiction. Moreover, a recent bibliometric review has highlighted a decline in the growth trend for psychotherapies applied for substance use disorders [5]. Advances in experimental psychology have informed the dual-process theoretical model [6,7], which postulates that existing therapies only address the cognitive control processes, but not the underlying automatic, unconscious processes. The dual-process theoretical model suggests there are two common automatic processes occurring in individuals with addictive disorders: attention bias and approach bias. Attention bias refers to the preferential allocation of attentional processes toward substance-related cues [8], whereas approach bias refers to the automated tendencies for individuals to seek out and reach for substance-related stimuli [9]. To assess these unconscious biases, tasks like the Stroop task and the Visual Probe task are commonly used [10]. These tasks are also used for bias modification, which focuses on retraining the biases, so that attention is directed away from the stimulus of interest [10].

The effectiveness of bias modification has been well studied. Cristea et al [11] reported moderate effectiveness (Hedges G=0.60) of cognitive bias modification in people with alcohol and tobacco use disorders. Other studies in clinical settings have similarly reported that bias modification helps reduce biases and other positive outcomes [9,12]. Most interventions included in these reviews were delivered in a laboratory setting [13], but with recent advances in technology, it is now possible to work outside the laboratory with Web-based and mobile-based interventions and provide early interventions and psychoeducation for addictive disorders [14]. Our recent review synthesized evidence for mobile attention bias interventions [15] and highlighted seven studies demonstrating effectiveness. Use of gamification technologies in these mobile interventions is the next advancement, as gamification might help diminish the repetitiveness of the task and increase both the extrinsic and intrinsic motivation to train [16]. Interestingly, in a review on gamified attention bias apps, only half of the studies adopting a gaming approach were more effective than standard care [17].

These inconsistent findings of the impact of gaming on the effectiveness of mobile interventions calls for further work to better understand the needs of patients (users) and health care professionals (providers). Participatory action research methods are a “systematic inquiry, with the participation of those affected by the problem being studied, for education and action or effecting social change” [18] and are well suited to the development of relevant interventions for the end user. Workshops and focus groups are the most common methods of cocreation, and these techniques are increasingly used in medicine. In our review of how participatory action methods have been applied for technological interventions in psychiatry [18], we reported seven studies describing how these methods have been used in the fields of perinatal depression, dementia, self-harm, and general mental health or youth mental health issues [18]. Such methods help in exploring perceptions and refining the existing task; they could also help researchers understand the reasons underlying the diminishing motivation and interest in such interventions [18]. The prior review focusing on participatory action research [18] highlighted the need to apply these methods for bias modification intervention research. Moreover, Zhang et al [19] also reported at least 17 bias modification apps in the commercial store, but only one app had an academic input. Similarly, in their review of smoking cessation apps, Haskin et al [20] found that only two of the validated apps were amongst the top 50 apps in the app store. It is evident from the review that there is a great divide between academics, developers, and the end users (or patients). Including methods of participatory design research will help bridge this disconnect between different stakeholders.

This study aimed to involve patients, together with health care professionals, in the design of a gamified mobile attention bias modification intervention. We wished to address four questions from the perspectives of health professionals and patients: (1) What are participant’s perspectives of the mobile attention bias modification intervention? (2) What features of a mobile attention bias modification intervention would a participant expect, to help minimize attrition from the task and increase both intrinsic and extrinsic motivation in completing the intervention? (3) Would gamification (ie, application of gaming elements) help in enhancing the existing mobile bias intervention task (ie, increase the magnitude of bias change, increase motivation, and reduce rates of attrition)? (4) What gaming elements are preferred?

Methods

Design Approach

The approach adopted was that of a Future Workshop, a method deemed to be optimal for the generation of new, innovative ideas [21]. We planned to conduct three co-design workshops: one with health care professionals, one with inpatients, and one with participants from outpatient settings.

Study Settings and Recruitment

Participants were recruited from the largest addiction treatment center in Singapore that specializes in both substance and behavioral addictions. The health care professionals recruited were addiction psychiatrists and addiction counsellors. The

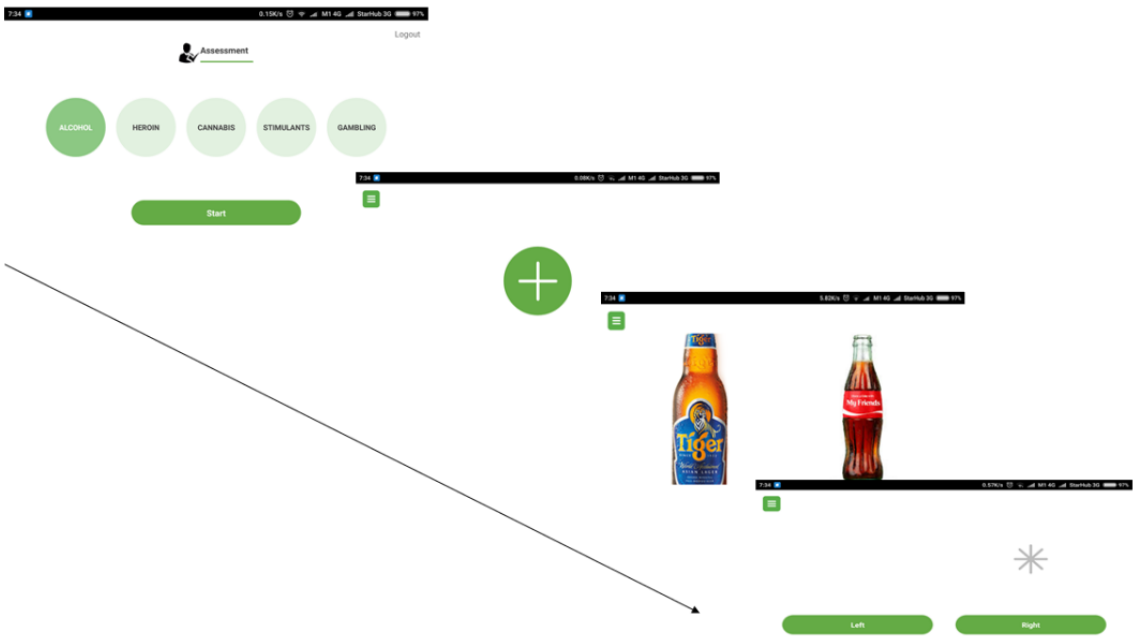
number of health care professionals and patients recruited was 10 for each group, so that there was equal representation from both groups of participants. In addition, patients who were at different stages of recovery were invited to participate, to ensure that the collated perspectives would not be biased. Informed consent was obtained in accordance with the Human Biomedical Research Guidelines from all participants.

Participants

Patients were included in the study if they were aged between 21 and 65 years; diagnosed with a primary psychiatric disorder of alcohol, opioid, cannabis, stimulant use disorder; able to speak and write in English; and knew how to use a smartphone device. Patients were excluded from the study if they had a significant psychiatric comorbidity (moderate to severe depressive disorder, anxiety disorder, psychotic disorder), were non–English-speaking, or had an existing cognitive impairment or intellectual disability. The inclusion criteria for selecting the patients were similar to those of our prior feasibility study [22] and representative of the demographics and clinical characteristics of patients admitted for treatment in our program.

For health care professionals to be included in the study, they had to be currently working in an addiction unit and actively involved in the treatment of individuals with addictive disorders with a minimum of 2 years of experience working with clients with addictive disorders.

Figure 1. Overview of the existing mobile attention bias intervention.



Participants were introduced to the rationale of the research project and the objectives of the study. Participants were then shown a presentation of the existing attention bias modification app. They were able to try the app on the provided tablet devices. Participants were asked to provide their thoughts narratively and identify limitations. The questions posed were as follows: (1) Having seen and used the existing app, what are your thoughts about it? (2) What are some of the limitations of the current app?

Workshop Procedures

All participants completed a questionnaire prior to participating in the co-design workshop comprising of three phases. The questionnaires included questions on the demographics of the health care professionals and demographics and clinical characteristics of the patients. The principal investigator (MZ), a psychiatrist specializing in addiction medicine, facilitated all three workshops.

Phase 1

In the first phase of the co-design workshop, individuals critiqued the existing smartphone-based attention bias modification intervention. The existing attention bias modification intervention was based on a prior protocol [23]. In the mobile app, participants were able to select the intervention specific for their substance of abuse. In the assessment task, participants were presented with a central fixation cross. Following the disappearance of the cross, both a neutral stimulus and a nonneutral stimulus were presented simultaneously. A probe would then replace either of the stimuli, and individuals were to indicate the position of the probe, within a predetermined response time. In the assessment task, the probe would replace either stimulus equally. In the intervention task, the probe would replace the nondrug stimulus all the time. Figure 1 provides an overview of the nature of the assessment and intervention tasks.

Phase 2

In the second phase, individuals brainstormed for solutions to their critiques. Participants were asked to brainstorm features that could be added to the existing mobile app. These questions were as follows: (1) What additional features do you think could

be added to the app? (2) Why do you think that these additional features will be helpful? Participants were encouraged to write down their ideas on sticky notes, following which these notes were collated by members of the study team and discussed amongst the whole group.

As one of the objectives was to determine whether gamification could address some of the limitations of the existing app, participants were shown screenshot examples of some of the attention bias commercial apps, which included elements of gamification [19]. Gamification elements, ie, digital rewards, avatar and competition, feedback, leader board, time pressure, and 3D environment, were shown to participants. To ensure everyone had some understanding of gamification techniques, the facilitator provided an overview of techniques summarized in the literature (Table 1) [24]. Participants were shown gaming strategies adopted in commercial apps and approaches used for apps in other disciplines and asked to consider strategies that are appropriate for the current intervention.

Table 1. Overview of gamification approaches [24].

Gaming approach	Description
Economic gamification techniques	
Marketplace and economies	Providing gamers with a virtual currency that allows them to deal in game
Digital rewards	These include badges, game currency, game points, virtual goods, and powers or abilities
Real-world prizes	Provides gamers with options to exchange in-game credits for real-world prizes such as vouchers or other forms of goods and services
Social gamification techniques	
Avatar	Allows individuals to choose a virtual character to represent oneself
Agent	A virtual character that help guides or provides instructions to user
Competition	Allows individuals to compete with other players or with each other
Teams	Game that involves several individual players, allowing them to interact and form relationships
Parallel communication systems	Allows individuals to communicate with one another
Social pressure	Ability of game to pressurize individuals to perform in certain task, so that he or she will be invited to subsequent events.
Performance-orientated techniques	
Feedback	Spoken, visual, or auditory feedback about user's performance
Levels	Information on the stage of a game one has attained
Secondary game objectives	Secondary goals that reward the player upon completion
Ranks of achievement	Measurement of character development
Leaderboards	Allows for comparisons with other players
Time pressure	Pre-determined time limits for task completion
Embedding-focused techniques	
Narrative context	A storyboard or stories that guide development of the character
3D environment	3D models of objects that parallel the real world

Participants were asked to discuss their perspectives about the gamification ideas described and then individually select their top three gamification techniques that they felt were most appropriate to be applied in the existing app.

Phase 3

In the last stage of the workshop, participants were divided into smaller groups of two to five to develop frame-by-frame sketches of a prototype app that incorporated the solutions that they have proposed. Participants sketched freely on the paper provided by using a variety of writing instruments provided. They were told to include the original task but could modify it. All the groups were given 15 minutes to work on this task.

Prior to the completion of the workshop, participants were also shown a set of substance images that have been incorporated since into the existing app. Participants were assigned to rate the relevance of the images for an individual with an addictive disorder, on a Likert scale (scores ranging from 1 to 10, where 1 is not relevant and 10 is extremely relevant). Participants were also asked to share their perceptions with the facilitator of the workshop.

Ethical Approval

This study obtained ethical approval from the National Healthcare Group's Domain Specific Research Board (ethics approval number 2018/01363).

Data Analyses

Descriptive statistical analysis was performed using SPSS (version 24; IBM Corp, Armonk, New York) for the quantitative data collated from the questionnaire. The workshops were

audio-recorded and transcribed verbatim by the principal investigator and independently by a private transcription service (WayWithWords). The principal investigator (MZ) then listened to the audio recordings of the workshop and developed a coding frame. To ensure reliability of the coding frame that was adopted, two authors (MZ and SH) reviewed the transcripts and discussed the coding frame, thus ensuring that the process of intercoder consensus was adhered to [25]. If there were any disagreements, they were resolved through discussion with another author. The codes identified were classified into categories and then reorganized into themes. NVivo (version 12.0; QSR International, Melbourne, Australia) was used to facilitate the analysis. Two independent members of the study team reviewed the sketches of the prototype and identified the common elements.

Results

Demographics of Participants

Three co-design workshops were conducted with health care professionals, inpatients, and outpatients. There were 20 participants, consisting of 10 health care professionals and 10 patients (5 inpatients on a detoxification and rehabilitation program and 5 outpatients who were clinically stable and mostly in abstinence from their substance use). The mean age of the health care professionals was 47.3 years, and the mean time caring for individuals with addictive disorders was 12.7 years. Of the 10 health care professionals, 8 were male. In terms of ethnicity, four were Chinese, one was Malay, and five were Indians. The mean age of inpatients was 44.4 years; one had an alcohol use disorder and four had a substance use disorder. For outpatients, the mean age was 43.2 years, and all participants had used both alcohol and illicit substances, except one inpatient. Multimedia Appendix 1 provides further information on the demographic characteristics of patient participants.

Phase 1

When responding to the existing app, health care professionals identified issues with its design, the visual probe trials, and the images. Both patient groups perceived issues with the visual probe trials and the images included in the app, and the outpatients also commented on the design and safety issues with the administration of the existing app. Table 2 provides a summary of the verbatim comments of the participants in each of the themes identified.

Table 2. Themes related to limitations of the existing app.

Themes	Health care professionals (n=10)	Inpatient participants (n=5)	Outpatient participants (n=5)
Design of the app	• "Buttons were too small" making them "easy to miss" [Participant 2] • "Press a few times, will add on to the reaction time" [Participant 1] • "There is no try out to understand how it works." [Participants 3 and 5]	• No mention	• "More instructions"[Participant 2] • "When I got things clear, I focus on star" [Participant 5] • "No design or anything" [Participant 1] • Buttons to be "bigger" [Participant 1] • "Joystick interface" to indicate a response [Participant 2]
Visual probe tasks	• "Too many repetitions" [Participant 2] • The task was "so fast" [Participants 3 and 5]. • "Distinguish the frame from one another looking at the asterisk" [Participant 1] • "Focusing on the asterisk only." [Participants 3 and 7] • "Pictures were so fast, after a while, I stopped paying attention and just look for the asterisk" [Participant 8]	• "Were too fast" [Participants 3, 5 and 7] • "My mind cannot catch up, fingers cannot catch up" [Participant 5] • "Repetition is the same" [Participant 5] • Task is "too tedious" [Participant 3]	• The trial was "too long" [Participants 1 and 4] • "Too fast" and "quick" [Participants 2, 4, and 5] • "I never even have a chance to see" [Participant 1] • "Too fast. My mind was to aim for the star. The pictures I was not interested." [Participant 4] • "It is too fast that we don't really see the picture" [Participant 3]
Images included in the visual probe task	• "Image colour does not stick out" [Participant 3] • "Non-white background will enhance the focus of the images" [Participant 2]	• The "colour is dull" and "the images are repetitive" [Participant 1] • "I see the same image for 30 times or more" [Participant 1]	• "Keep seeing the same pictures" [Participant 3] • "Very boring, put more pictures" [Participant 1] • "Maybe more picture, very repetitive. I keep on seeing the same pictures over and over again" [Participant 2] • "It comes in pairs. It always weed and chocolate cake. You can figure out" [Participant 2]
Safety of administering the app	• No mention	• No mention	• "Triggering. Like I think just keep seeing pictures of drug of choice" [Participant 2] • "Early recovery is triggering. For myself, if I am in detox, might trigger me." [Participant 5]

Phase 2

When brainstorming features that could be added to the existing mobile app, the health care professionals suggested improvement to the stimuli, the mechanism of responding, and the presentation of the scores. Inpatient participants recommended additional functionalities and enhancements in the design, images, and task. Outpatients recommended improvements in both the design, included images, and the presentation of the results and perceived gaming features to be a solution. Table 3 provides a summary of the verbatim comments of the participants in accordance with the themes identified.

Table 3. Themes related to solutions addressing limitations of the existing app.

Themes	Health care professionals (n=10)	Inpatient participants (n=5)	Outpatient participants (n=5)
Images included in visual probe task	• "As much as possible, it should be similar as possible. The two pictures should be of the same quality and same size" [Participant 2] • "How come the tiger bottle is so big and the cola bottle so small? Obviously, my eyes will zoom to the big one. I am attracted to the obvious. It just shines out" [Participant 5]	• "Something associated with alcoholism. Clubbing, coffee shop, anything associated with alcoholism" [Participant 1] • Images with substances in different "environment" [Participant 3] • "Can include like family members. Drinking with family members" [Participant 2] • "Images personalized, can identify with" [Participant 3]	• "Higher intensity" of images" • instead of just 2 pictures [Participant 1]
Design of app	• "Is it possible to press on the picture instead of right/left. Pictures are bigger than the right/left button" [Participant 10] • "Maybe instead of buttons, they could tap anywhere on the half of the screen. That would make it easier rather than to aim on the button. This will also help to mitigate the older folks who have difficulties to move on to the button" [Participant 2]	• "Colours" for the asterisk [Participant 1] and for the buttons to be "round or bigger" [Participant 1]	• "When press down on the star, right or correct, some positive words or pictures, smiling. Would be more interesting" [Participant 4] • "If they hit the correct one, maybe there is a nice emoji. Wrong, maybe a crying emoji" [Participant 3]
Scoring functionality	• Scoring was "very complicated" and that the scoring appealed only to "investigators or clinicians" [Participant 1]	• No mention	• "When they finish the game, can have instant results, instead of having to scroll down" [Participant 3]
Other possible functionalities	• No mention	• "What are the diseases you are going through if you have alcoholism? More information of what alcohol does to you in the short term and long term" [Participant 2]	• No mention
Visual probe trial	• No mention	• "Just 3 seconds interval" [Participant 1] • "You can request for 5-10 seconds interval delay" [Participant 1]	• "Do it from slow, all the way to fast" [Participant 3]
Consideration of gaming elements	• No mention	• No mention	• "Show the fastest speed and slowest speech in the result, make it like a game" [Participant 1] • "Maybe you could have a board [1st, 2nd, 3rd)" [Participant 3]

Phase 3

Gaming was the aspect where there was the greatest difference between health care professionals and patients. The professionals suggested the integration into the conventional attention bias modification app of performance-oriented rewards and storyboard gaming elements. With one exception, the inpatient participants were against the inclusion of gaming elements. The outpatient participants perceived that the inclusion of gaming elements was appropriate if the game was intended for individuals who are in abstinence. Table 4 provides a summary of the verbatim comments of the participants in accordance to the themes identified.

Table 4. Themes related to gamification elements for the existing app.

Themes	Health care professionals (n=10)	Inpatient participants (n=5)	Outpatient participants (n=5)
Performance-oriented gaming elements	• "Lowest cost way of doing this app that will make users want to continue using it." [Participant 2] • "The feedback that you are doing better is kind of helpful for them" [Participant 9] • "In between giving feedback about their performance; three times you press the correct thing, you are on a roll" [Participant 4] • "As it motivates you to go into higher grades, either as an individual or as a group" [Participant 6] • "Motivate them to continue playing the application" [Participant 11] • "Bronze to silver to gold medals" [Participant 2] • "Some achievement in the form of badges would help" [Participant 7] • "Track their progress" "aware of the questions remaining" [Participant 8]	• "Leader-board", social gaming elements "Having connect with Facebook would allow you to see who else has participated" and "Levels" in the game [Participant 1]	• "Results" of his or her performance [Participant 2] • Time-pressure, levels and leaderboard were also chosen, as it allows for "competition" [Participant 2] • "When you see the leaderboard, maybe I am here. I can do better to go up" [Participant 3]
Rewards (digital/real world)-based gaming elements	• "Entice them to continue to play" [Participant 4]	• No mention	• "Can change for vouchers" [Participant 1] • "Like voucher, exchange coffee. Get something realistic" [Participant 4]
Context (storyboard)	• "Motivating" [Participant 7] • "Map out the real-life experiences", this giving "a reason in doing the exercises" [Participant 6] • "More realistic and engaging" [Participant 9]	• No mention	• No mention
Against gamification	• No mention	• "Childish" [Participant 3] • making the application to be more catered for "Kids" [Participant 1] • "All these are game right? Part of games right? All these are addiction already" [Participant 2] • "As a gamer, I could relate to Number 2, this is gaming addiction" [Participant 1]	• "If I am using, I would not play game. When you are into drugs, you are in no mood to play games" [Participant 4] • "When you are heavily using, you cannot be bothered by this. Especially when you are high" [Participant 5]

Each of the six groups were able to discuss and sketch out a prototype of an app that would address the underlying limitations and include the solutions they have proposed. Figure 2 shows an example of the proposed design illustrated by one group of patients. Table 5 provides the selected verbatim comments of all groups of participants with regard to their sketched prototype. In each of the groups, one participant was asked to share about their prototype.

Figure 2. Patient-created proposal for design of the app.

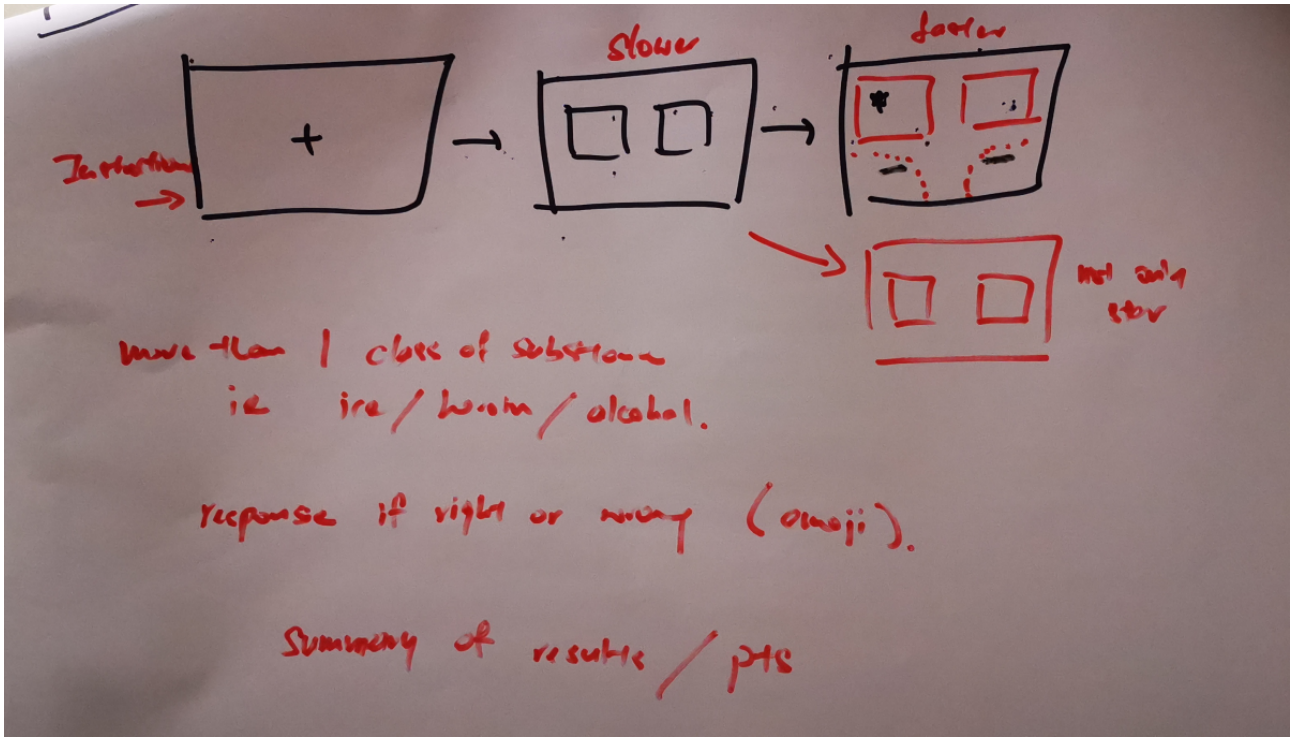


Table 5. Participants' comments on their newly conceptualized prototypes (one participant was asked to share on behalf of the whole group).

Participants	Selected verbatim comments
Health care professionals (Group 1)	• "We think that there should be a few levels of difficulties. It is like a training phase. The assessment time becomes shorter and shorter. The time becomes shorter and shorter. We think there should be a score. Encourage the person to hit a higher score. The rewards we were thinking about…Vouchers or privileges on the ward."
Health care Professionals (Group 2)	• "First thing would be instructions." • "Second part. There would be some motivation. You are doing good." • "After the training is done, task score. And the progress bar. How many percent completed." • "The pictures. We thought of different modalities. Not just the pictures. Sometimes virtual ones." • "Need a trial of how to do it."
Inpatient participants (Group 1)	• "Instead of using a cross, you can put a money sign." • "After you focus on a while, you could have different images." • "Secondly, instead of money, you could put a photo. A photo of your family. After you focus on the photo for a while, you drift off to an animated picture." • "Ask the user to type."
Inpatient participants (Group 2)	• "Family. Picture This is the family picture, wife and children picture. Drugs and alcohol. Which one you want to choose." • "Instead of someone else drinking, can be someone I know drinking."
Outpatient participants (Group 1)	• "The first square is the instruction manual. The second one is a prompt for ready. Then straight away go to the cross with 2 pictures. The last one is the start. And end with results. Maybe it starts off with a bit slower and then progress to fast." • "Maybe the first 20-30 pictures 0.8 seconds, then the next 0.6 seconds. Moderately decrease the speed instead of diving right in."
Outpatient participants (Group 2)	• "Introduction. The second column slightly slower than the tablet. 0.7-0.8 seconds. This one is 2 seconds. 1 seconds. Then you can random got different picture. Not only just come up star only. Can come up star with drug."

In both groups of health care professionals, it was important that participants be allowed to have a training phase or practice before they undertake the actual task. Health care professionals also recommended the initial trial interval to be lengthened to 1000 milliseconds and gradually reduced to 500 milliseconds with time. Health care professionals also recommended the

inclusion of scores and progress bars, to provide feedback and inform participants of their progress. Some health care professionals preferred the inclusion of animated images instead of static images. Inpatients wanted images that were animated or personalized for their substance of use. Participants also wanted a variety of images to appear on the screen, instead of being limited to merely two images. Like the design proposed by health care professionals, both groups of outpatient participants wanted to start with a lengthened trial interval, which was progressively shortened as the intervention progressed, for example, starting with a trial interval of 800 milliseconds for the first 20-30 set of images and then reducing it to 600 milliseconds. The rationale for this progressively decreasing trial interval duration was to allow participants of all age groups to become comfortable with the nature of the task. All participants emphasized the need to provide instructions prior to the commencement of the task. Both groups of participants also wanted the prototype to have bigger buttons to facilitate rapid response. All participants were able to complete the sketches of the prototype within the allocated time.

Participants were also shown images that were included in the existing app. Both health care professionals and participants shared their perspectives of the included images. Table 6 provides a summary of the verbatim comments of the participants in accordance with the drug categories.

Table 6. Suggestions of participants for images included in the app.

Image category	Health care professionals (n=10)	Inpatient participants (n=5)	Outpatient participants (n=5)
Alcohol	• "Picture more relevant to the local context, for example, picture from coffee shop" [Participant 6] • Not to focus on the "brands" of the alcohol [Participant 6]	• "Hard liquor" was missing [Participants 1 and 2]	• No need to include the "brand" of the alcohol [Participant 4] • "More kind of hard liquor, such as whisky" [Participant 4]
Cannabis	• "Some of the pictures were not familiar" [Participant 2]	• A "barrel and aluminum foil" would be good [Participant 3] • "Bottles" might not be so relevant, as they are synthetic cannabinoids [Participant 1]	• "The bong is not the right bong" [Participant 2]
Heroin	• "Straws" "pictures of the barrel and thin foils" [Participant 5] • "Someone chasing the dragon, or someone using the main line" are the most relevant and triggering [Participant 3]	• "The one with needle" [Participant 4]	• "Syringe" image was perceived to be "very triggering" [Participant 1] • "Nothing catches my eye except for this needle thing" [Participant 3] • Pills "oxycodone" was not used locally ("Singapore people don't abuse"). [Participant 1]
Stimulants	• The images of the stimulant crystals might not be so relevant, as the crystals available locally are "of different quality" [Participant 2]	• "Pure image of ICE[a]" [Participant 1]	• "Don't look like ICE at all" [Participant 4]

[a]ICE: methamphetamine (stimulant).

Discussion

Principal Findings

This is the first study to explore the principles of participatory design research for cognitive/attention bias modification interventions. When asked to identify the limitations in the existing app, common issues identified were those with the design, visual probe task, and included images. Outpatients were also concerned with the safety of administration of the intervention. In the brainstorming sessions, health care professionals made recommendations for enhancing the stimulus, mechanism of responding, and presentation of scores. Inpatient participants recommended the addition of functionalities such as information on the harms associated with the substance use and enhancements in the design, images, and tasks. Outpatient participants perceived the need to improve the images and presentation of the results and recommended the inclusion of gaming features. There were differences in opinions pertaining to the inclusion of gaming features, as only health care professionals endorsed their inclusion. In the last phase of the workshop, participants were tasked with the conceptualization of prototypes, and the commonality in the design was a gradual shortening of the interval for stimulus/image presentation.

Throughout all the co-design sessions, one of the main issues highlighted by participants was that the time interval for the presentation of the stimulus was too rapid. In the conventional app that we designed [23], we stipulated the following timings: 500 milliseconds for the presentation of the fixation cross, another 500 milliseconds for the presentation of the images, and 2000 milliseconds for the participant to respond before the trial goes on. A measure of 500 milliseconds was chosen as the time for the presentation of the stimulus, as most studies that have examined the reliability of the dot-probe task present cue stimuli 500 milliseconds prior to the appearance of the probe [26]. In the literature examining attentional bias modification for substance use disorder, there is a great variation in the

timings of stimulus presentation. Charles et al [27] presented images for either 200 or 500 milliseconds to individuals with opioid use disorders, and they postulate that the short stimulus (200 milliseconds) helped in the evaluation of the automatic orientating, and the long stimulus (500 milliseconds) helped in the evaluation of controlled attentional processing. Other studies involving individuals with opioid disorders have used varied timings, ranging from a short stimulus of 200 milliseconds [28-32] to 500 milliseconds [29] and a long stimulus interval of 1500 milliseconds [15] to 2000 milliseconds [28,30-32]. In studies involving individuals with cannabis use disorder, stimulus intervals of 500 milliseconds [33] and 2000 milliseconds [34] have also been used. Studies involving participants with stimulant use disorders have, however, consistently used a stimulus interval of 500 milliseconds [35-37]. The aspects postulated by Charles et al [27] in their study on the evaluation of different attentional processes by short and long stimulus intervals have also been previously postulated by Robbins and Ehrman [38] and Field et al [39]. It is apparent that there is a wide variability in the stimulus interval timings in the published literature, despite all these studies having used the visual probe task. The findings from this study suggest that participants (both health care professionals and patients) prefer a slightly lengthened stimulus presentation interval (700-1000 milliseconds) and with the stimulus presentation interval gradually decreasing across the interventions. Taking into consideration the perspectives of our participants and the evidence to date in the published studies, future studies could vary the ratio of the stimuli that present as a long stimulus and a short stimulus (ie, at the start of the intervention, more trials are presented for a long stimulus duration, and across the days of interventions, this progressively decreases, so that there are more trials presented for a short stimulus duration). This helps ensure that both attentional processes (initial orientating and delayed disengagement) are being investigated.

One of the other limitations that were consistently highlighted throughout the workshops was a need for the stimulus images to be more relevant and personalized. The images included in the existing app were images extracted from the United States Drug Enforcement Agency Website together with some from the Singapore's Central Narcotics Bureau's website. Participants suggested that some of these images are not realistic enough for them and are not congruent with the images of the substances they have had used. To bridge this limitation, we could consider what Field et al [33] previously adopted. They had a separate group of 10 individuals, comprised of cannabis and noncannabis users, to rate the word stimulus on a "cannabis-relatedness" scale. They then used words that had the highest rating scores. No recent studies have replicated this study, and we propose that this should be considered in future research to ensure that the images included are relatable to the participants. Apart from relevance and relatedness, personalization of the images was suggested by inpatient participants. Although Field et al [40] reported that the poor reliability of the visual probe task could be attributed to the nature of the stimulus used and highlighted the need for personalization of images, a recent study by Jones et al [41] failed to demonstrate increased internal consistency of the visual probe task following personalization. Their study [41], which involved participants with alcohol use issues, presented participants with images related to their type of alcoholic beverage consumed rather than a broad range of alcohol images. Although an attractive option, it is evident from the published literature that personalization of images may not improve the reliability of the visual probe task and facilitate the detection of attentional biases. Moreover, there are privacy and practical constrains if personalization of the images is considered.

In this study, we found a discrepancy in the perspectives of health care professionals and patients with regard to the consideration of gaming elements. Health care professionals were open to the inclusion of gaming elements, whereas patients were cautious in considering gaming elements. It is extremely important to take into consideration the needs of patients, as they are the eventual final users. Allowing patients' perspectives to take dominance over health care and academic perspective makes sense, given that it is increasingly recognized that patients bring unique insights and knowledge into the co-designing process [42]. This discrepancy in viewpoints could potentially be mitigated if health care professionals and patients are both included in the same co-design workshop. Unfortunately, this was not permitted by our ethical board, as there are concerns that patient participants might not be as vocal due to the presence of health care professionals. In our study, some patients acknowledged that gaming elements could enhance the existing task. The common gaming elements recommended by both health care professionals and patients were performance-oriented gaming elements (feedback, levels, etc) and rewards. In their review of gamified attention bias modification apps, Zhang et al [43] highlighted that apps previously evaluated included features like animations, sounds, feedback, and a point-scoring system for response time and difficulty. To some extent, our participants have endorsed similar gamification techniques, as the gaming elements that have been used previously could be easily clustered as performance-oriented gaming elements, according to a previous [24] taxonomy of gamification techniques. In the review by Zhang et al [43], only two of the four identified studies demonstrated that the gamified app was effective. Notably, the two studies that were efficacious were evaluating the same app. One study [44] that involved participants with alcohol use disorders did not find the game to be effective following the inclusion of gaming elements. Given our findings, it is pertinent for future studies to carefully consider the appropriate use of gaming elements, considering mainly the perspectives of patients, and to re-evaluate the app for its effectiveness. Future research should consider, in particular, the specific type of gaming strategy that would render the intervention more effective.

In our study, patients also highlighted the presence of safety concerns in the administration of such an attention bias modification app to individuals in the early stages of recovery. Their concerns are contrary to those of prior studies, which have introduced attention bias modification to individuals who were in the detoxification phase of their treatment [9,12]. Manning et al [12] highlighted that it is important to introduce bias modification early to capitalize on the stage of neural recovery. None of these prior studies have reported dropouts due to individuals relapsing into their substance of use. Nevertheless,

the safety concerns highlighted by patients should be taken into consideration when executing a trial. It might be more appropriate to consider the administration of such an intervention among individuals in the later stages of their rehabilitation program or when they are out of their withdrawal phase. If the intervention must be administered to participants in the detoxification phase, only participants with low scores on their withdrawal scales should be selected to undertake such an intervention.

The main strength of this study is the principles of participatory design research for the coproduction of an attention bias modification app. The coproduced app not only included the perspectives of experts in the field of addiction (health care professionals), but also took into consideration the needs of the service users. Most of the existing attention bias and cognitive bias modification interventions are produced by either academic or commercial developers, and as highlighted by Zhang et al [15], there remains an apparent disconnect between academics, health care professionals, and software developers. This study has clearly bridged this gap in the research literature. In this study, we conducted the workshop in line with the principles of the Future Workshops.

Despite these inherent strengths, there are several limitations. It would have been most ideal to have both health care professionals and participants in the same workshop. However, we were unable to do so, as there is a possibility that the presence of health care professionals might make it uncomfortable for participants to share their inputs and perspectives. From the perspectives shared by the participants and from the results, it is evident that not all participants had the same understanding of the visual probe task paradigm, despite receiving the same introductory briefing. In addition, some of the themes identified such as cost were of importance, but the cost was not a theme that was commonly identified by both health care professionals and patients; therefore, it was not further explored. In our study, we showed participants examples of the various gamification strategies, as it is anticipated that our participants had a limited understanding of these techniques. There is a possibility that this might have resulted in some biases among individuals, but we have also taken steps to minimize this by ensuring that gamification techniques from all four categories, as previously highlighted by Hoffman et al [24], are shown. In addition, to minimize the risk of biases, in all three workshops, the facilitator explained each of the gamification strategies listed in the prior classification by Hoffman et al [24].

Conclusions

To the best of our knowledge, this is the first study that has applied the principles of participatory design research for an attention bias modification intervention. Both health care professionals and patient participants provided insights on how the existing paradigm of the attention bias modification task could be enhanced to meet their needs. The results from this research will guide the development of an app that meets the specific needs of patients and is still based on a pre-existing validated task paradigm.

Acknowledgments

MZ is supported by a grant under the Singapore Ministry of Health's National Medical Research Council (grant number NMRC/Fellowship/0048/2017) for PhD training. This study has received funding from the National Healthcare Group & Lee Kong Chian School of Medicine, ALIVE (Games for Health Innovation Centre) grant.

Authors' Contributions

MZ, SH, GS, DF, and HS jointly conceptualized and designed the study. MZ and SH collated and analyzed the data. MZ wrote up the draft of the initial manuscript under the guidance of DF and HS. All authors worked on the final manuscript. All authors read and approved of the protocol prior to submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic characteristics of patient participants.
[PDF File (Adobe PDF File)91 KB - mhealth_v7i9e15871_app1.pdf]

References

1. United NOOD. Executive Summary: Conclusions & Policy Implementation. 2017 URL: https://www.unodc.org/wdr2017/field/Booklet_1_EXSUM.pdf [accessed 2019-08-14]
2. Ho R, Ho E, Mak A. Cutaneous complications among i.v. buprenorphine users. J Dermatol 2009 Jan;36(1):22-29. [doi: 10.1111/j.1346-8138.2008.00581.x] [Medline: 19207433]
3. Ho RCM, Ho ECL, Tan CH, Mak A. Pulmonary hypertension in first episode infective endocarditis among intravenous buprenorphine users: case report. Am J Drug Alcohol Abuse 2009 Jul 21;35(3):199-202. [doi: 10.1080/00952990902939719] [Medline: 19462305]

4. Cutler RB, Fishbain DA. Are alcoholism treatments effective? The Project MATCH data. BMC Public Health 2005 Jul 14;5(1). [doi: 10.1186/1471-2458-5-75]
5. Tran BX, Moir M, Latkin CA, Hall BJ, Nguyen CT, Ha GH, et al. Global research mapping of substance use disorder and treatment 1971-2017: implications for priority setting. Subst Abuse Treat Prev Policy 2019 May 17;14(1):21 [FREE Full text] [doi: 10.1186/s13011-019-0204-7] [Medline: 31101059]
6. Field M, Cox WM. Attentional bias in addictive behaviors: a review of its development, causes, and consequences. Drug Alcohol Depend 2008 Sep 01;97(1-2):1-20. [doi: 10.1016/j.drugalcdep.2008.03.030] [Medline: 18479844]
7. Stacy AW, Wiers RW. Implicit cognition and addiction: a tool for explaining paradoxical behavior. Annu Rev Clin Psychol 2010 Mar;6(1):551-575 [FREE Full text] [doi: 10.1146/annurev.clinpsy.121208.131444] [Medline: 20192786]
8. Field M, Marhe R, Franken IHA. The clinical relevance of attentional bias in substance use disorders. CNS Spectr 2013 May 13;19(3):225-230. [doi: 10.1017/s1092852913000321]
9. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013 Apr;4:38-51 [FREE Full text] [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
10. Zhang M, Ying J, Song G, Fung D, Smith H. Recent Advances in Attention Bias Modification for Substance Addictions. IJERPH 2018 Apr 04;15(4):676. [doi: 10.3390/ijerph15040676]
11. Cristea IA, Kok RN, Cuijpers P. The Effectiveness of Cognitive Bias Modification Interventions for Substance Addictions: A Meta-Analysis. PLoS One 2016 Sep;11(9):e0162226 [FREE Full text] [doi: 10.1371/journal.pone.0162226] [Medline: 27611692]
12. Manning V, Staiger P, Hall K, Garfield J, Flaks G, Leung D, et al. Cognitive Bias Modification Training During Inpatient Alcohol Detoxification Reduces Early Relapse: A Randomized Controlled Trial. Alcohol Clin Exp Res 2016 Sep;40(9):2011-2019. [doi: 10.1111/acer.13163] [Medline: 27488392]
13. Jones EB, Sharpe L. Cognitive bias modification: A review of meta-analyses. J Affect Disord 2017 Dec 01;223:175-183. [doi: 10.1016/j.jad.2017.07.034] [Medline: 28759865]
14. Zhang MWB, Ho RCM. Tapping onto the Potential of Smartphone Applications for Psycho-Education and Early Intervention in Addictions. Front Psychiatry 2016 Mar;7:40 [FREE Full text] [doi: 10.3389/fpsyt.2016.00040] [Medline: 27014103]
15. Zhang M, Ying J, Song G, Fung DS, Smith H. Attention and Cognitive Bias Modification Apps: Review of the Literature and of Commercially Available Apps. JMIR Mhealth Uhealth 2018 May 24;6(5):e10034 [FREE Full text] [doi: 10.2196/10034] [Medline: 29793899]
16. Dalstrom C. Impacts of gamification on Intrinsic Motivation. URL: https://www.ntnu.edu/documents/139799/1279149990/04+Article+Final_camildah_fors%C3%B8k_2017-12-06-13-53-55_TPD4505.Camilla.Dahlstr%C3%B8m.pdf/9e48c5f5-0d17-4276-a23e-434abfe65491, [accessed 2003-09-19]
17. Zhang M, Ying J, Song G, Fung DS, Smith H. Gamified Cognitive Bias Modification Interventions for Psychiatric Disorders: Review. JMIR Ment Health 2018 Oct 25;5(4):e11640. [doi: 10.2196/11640]
18. Zhang M, Ying J. Incorporating Participatory Action Research in Attention Bias Modification Interventions for Addictive Disorders: Perspectives. IJERPH 2019 Mar 06;16(5):822. [doi: 10.3390/ijerph16050822]
19. Zhang MW, Ying JB, Song G, Ho RC. A review of gamification approaches in commercial cognitive bias modification gaming applications. THC 2018 Dec 06;26(6):933-944. [doi: 10.3233/thc-181313]
20. 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]
21. Jungk, Robert, Müllert, Norbert. Future workshops: How to Create Desirable Futures. London, England: Institute for Social Inventions; 1987.
22. Zhang M, Ying J, Amron S, Mahreen Z, Song G, Fung D, et al. A Smartphone Attention Bias App for Individuals With Addictive Disorders: Feasibility and Acceptability Study. JMIR Mhealth Uhealth 2019 Sep 13;7(9):e15465 [FREE Full text] [doi: 10.2196/15465]
23. Zhang M, Ying J, Song G, Fung DS, Smith H. Mobile Phone Cognitive Bias Modification Research Platform for Substance Use Disorders: Protocol for a Feasibility Study. JMIR Res Protoc 2018 Jun 12;7(6):e153. [doi: 10.2196/resprot.9740]
24. Hoffmann A, Christmann CA, Bleser G. Gamification in Stress Management Apps: A Critical App Review. JMIR Serious Games 2017 Jun 07;5(2):e13. [doi: 10.2196/games.7216]
25. Schielke HJ, Fishman JL, Osatuke K, Stiles WB. Creative consensus on interpretations of qualitative data: the Ward method. Psychother Res 2009 Jul 22;19(4-5):558-565. [doi: 10.1080/10503300802621180] [Medline: 19235093]
26. Chapman A, Devue C, Grimshaw GM. Fleeting reliability in the dot-probe task. Psychological Research 2017 Nov 20;83(2):308-320. [doi: 10.1007/s00426-017-0947-6]
27. Charles M, Wellington C, Mokrysz C, Freeman T, O'Ryan D, Curran H. Attentional bias and treatment adherence in substitute-prescribed opiate users. Addict Behav 2015 Jul;46:100-105. [doi: 10.1016/j.addbeh.2015.03.017] [Medline: 25838001]

28. Constantinou N, Morgan CJ, Battistella S, O'Ryan D, Davis P, Curran HV. Attentional bias, inhibitory control and acute stress in current and former opiate addicts. Drug Alcohol Depend 2010 Jun 01;109(1-3):220-225. [doi: 10.1016/j.drugalcdep.2010.01.012] [Medline: 20172662]
29. Frankland L, Bradley BP, Mogg K. Time course of attentional bias to drug cues in opioid dependence. Psychol Addict Behav 2016 Aug;30(5):601-606. [doi: 10.1037/adb0000169] [Medline: 27031089]
30. Garland EL, Froeliger B, Howard MO. Allostatic dysregulation of natural reward processing in prescription opioid misuse: autonomic and attentional evidence. Biol Psychol 2015 Feb;105:124-129 [FREE Full text] [doi: 10.1016/j.biopsycho.2015.01.005] [Medline: 25603284]
31. Garland EL, Froeliger BE, Passik SD, Howard MO. Attentional bias for prescription opioid cues among opioid dependent chronic pain patients. J Behav Med 2013 Dec 12;36(6):611-620 [FREE Full text] [doi: 10.1007/s10865-012-9455-8] [Medline: 22968666]
32. Garland EL, Howard MO. Opioid attentional bias and cue-elicited craving predict future risk of prescription opioid misuse among chronic pain patients. Drug Alcohol Depend 2014 Nov 01;144:283-287 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.09.014] [Medline: 25282309]
33. Field M, Mogg K, Bradley BP. Cognitive bias and drug craving in recreational cannabis users. Drug Alcohol Depend 2004 Apr 09;74(1):105-111. [doi: 10.1016/j.drugalcdep.2003.12.005] [Medline: 15072814]
34. Field M, Eastwood B, Bradley BP, Mogg K. Selective processing of cannabis cues in regular cannabis users. Drug Alcohol Depend 2006 Oct 15;85(1):75-82. [doi: 10.1016/j.drugalcdep.2006.03.018] [Medline: 16701963]
35. Montgomery C, Field M, Atkinson AM, Cole JC, Goudie AJ, Sumnall HR. Effects of alcohol preload on attentional bias towards cocaine-related cues. Psychopharmacology (Berl) 2010 Jun 30;210(3):365-375. [doi: 10.1007/s00213-010-1830-y] [Medline: 20352412]
36. Tull MT, McDermott MJ, Gratz KL, Coffey SF, Lejuez CW. Cocaine-related attentional bias following trauma cue exposure among cocaine dependent in-patients with and without post-traumatic stress disorder. Addiction 2011 Oct;106(10):1810-1818 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03508.x] [Medline: 21615582]
37. Bardeen JR, Dixon-Gordon KL, Tull MT, Lyons JA, Gratz KL. An investigation of the relationship between borderline personality disorder and cocaine-related attentional bias following trauma cue exposure: the moderating role of gender. Compr Psychiatry 2014 Jan;55(1):113-122 [FREE Full text] [doi: 10.1016/j.comppsych.2013.08.011] [Medline: 24138957]
38. Robbins SJ, Ehrman RN. The role of attentional bias in substance abuse. Behav Cogn Neurosci Rev 2004 Dec 18;3(4):243-260. [doi: 10.1177/1534582305275423] [Medline: 15812109]
39. Field M, Mogg K, Zetteler J, Bradley BP. Attentional biases for alcohol cues in heavy and light social drinkers: the roles of initial orienting and maintained attention. Psychopharmacology (Berl) 2004 Oct 8;176(1):88-93. [doi: 10.1007/s00213-004-1855-1] [Medline: 15071718]
40. Field M, Christiansen P. Commentary on, 'Internal reliability of measures of substance-related cognitive bias'. Drug Alcohol Depend 2012 Aug 01;124(3):189-190. [doi: 10.1016/j.drugalcdep.2012.02.009] [Medline: 22386685]
41. Jones A, Christiansen P, Field M. Failed attempts to improve the reliability of the alcohol visual probe task following empirical recommendations. Psychol Addict Behav 2018 Dec;32(8):922-932 [FREE Full text] [doi: 10.1037/adb0000414] [Medline: 30475013]
42. Manzini E, Rizzo F. Small projects/large changes: Participatory design as an open participated process. CoDesign 2011 Sep;7(3-4):199-215. [doi: 10.1080/15710882.2011.630472]
43. Zhang M, Ying J, Ho RC. Gamified Cognitive Bias Interventions for Psychiatric Disorders: Protocol of a Systematic Review. JMIR Res Protoc 2018 Oct 16;7(10):e10154. [doi: 10.2196/10154]
44. Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW. Attentional Bias Modification With Serious Game Elements: Evaluating the Shots Game. JMIR Serious Games 2016 Dec 06;4(2):e20. [doi: 10.2196/games.6464]

Edited by G Eysenbach; submitted 14.08.19; peer-reviewed by R Ho, L Notebaert, C Magnusson, K Blondon; comments to author 31.08.19; revised version received 03.09.19; accepted 05.09.19; published 03.10.19.

Please cite as:
Zhang M, Heng S, Song G, Fung DSS, Smith HE
Co-designing a Mobile Gamified Attention Bias Modification Intervention for Substance Use Disorders: Participatory Research Study
JMIR Mhealth Uhealth 2019;7(10):e15871
URL: https://mhealth.jmir.org/2019/10/e15871
doi:10.2196/15871
PMID:31584003

Original Paper

Development of an Ambulatory Biofeedback App to Enhance Emotional Awareness in Patients with Borderline Personality Disorder: Multicycle Usability Testing Study

Youri PMJ Derks[1,2], MSc; Randy Klaassen[3], PhD; Gerben J Westerhof[1], PhD; Ernst T Bohlmeijer[1], PhD; Matthijs L Noordzij[1], PhD

[1]Department of Psychology, Health, and Technology, University of Twente, Enschede, Netherlands

[2]Scelta, Center for Treatment of Personality Disorders, GGNet, Mental Health Institute, Apeldoorn, Netherlands

[3]Department of Human Media Interaction, University of Twente, Enschede, Netherlands

Corresponding Author:
Youri PMJ Derks, MSc
Department of Psychology, Health, and Technology
University of Twente
Cubicus Bldg, 1st Fl
5, Drienerlolaan, 7522 NB Enschede
Enschede, 7500 AE
Netherlands
Phone: 31 889335253
Email: y.p.m.j.derks@utwente.nl

Abstract

Background: Patients with borderline personality disorder experience great difficulties in regulating their emotions. They often are unable to effectively detect their emotional arousal and struggle to timely apply learned techniques for emotion regulation. Although the use of continuous wearable biofeedback has been repeatedly suggested as an option to improve patients' emotional awareness, this type of app is not yet available for clinical use. Therefore, we developed an ambulatory biofeedback app named Sense-IT that can be integrated in mental health care.

Objective: The aim of the study was to develop an ambulatory biofeedback app for mental health care that helps with learning to better recognize changes in personal emotional arousal and increases emotional awareness.

Methods: Using several methods in a tailored User Centred Design (UCD) framework, we tested the app's usability and user experience (UX) via a cyclic developmental process with multiple user groups (patients, therapists, and UCD experts; 3-5 per group, per cycle).

Results: The process resulted in a stable prototype of the app that meets most of the identified user requirements. The app was valued as useful and usable by involved patients, therapists, and UCD experts. On the Subjective Usability Scale (SUS), the patients rated the app as "Good" (average score of 78.8), whereas the therapists rated the app as "OK" (average score of 59.4). The UCD experts judged the app's overall usability as between "OK" and "acceptable" (average score of 0.87 on a cognitive walkthrough). As most critical usability problems were identified and addressed in the first cycle of the prototyping process, subsequent cycles were mainly about implementing new or extending existing functions, and other adjustments to improve UX.

Conclusions: mHealth development within a clinical mental health setting is challenging, yet feasible and welcomed by targeted users. This paper shows how new mHealth interventions for mental health care can be met with enthusiasm and openness by user groups that are known to be reluctant to embrace technological innovations. The use of the UCD framework, involving multiple user groups, proved to be of added value during design and realization as evidenced by the complementary requirements and perspectives. Future directions on studying clinical effectiveness of the app, appliance of the app in other fields, and the implications of integration of the app for daily practice in mental health are discussed.

(JMIR Mhealth Uhealth 2019;7(10):e13479) doi:10.2196/13479

KEYWORDS

mental health; borderline personality disorder; mobile health; emotional awareness; ambulatory biofeedback; design science; user centered design; iterative prototype testing

Introduction

Context

Borderline personality disorder (BPD) is a psychological disorder that influences all domains of life. It is characterized by a pervasive pattern of unstable relations, a distorted self-image, and profound difficulties in regulation of one's emotions [1]. Self-harming behaviors are common [2-5]. A lack of emotional awareness or the ability to timely recognize the onset of emotions and their increasing or decreasing intensity plays a role in emotional instability and dysregulated and self-harming behaviors in BPD [6-11]. Patients with BPD seem to especially have less focus on the level of emotional arousal than controls [12].

Perception of internal bodily states is found to be of significant importance to subjective experience, awareness, labeling, and understanding of emotional processes [13-16]. Interestingly, people with low emotional awareness, in general, *do* respond to emotional triggers physiologically, and to a certain extent, behaviorally, but lack the experience of the emotion—the *feeling* [14,17,18]. Therefore, available evidence suggests that treatment of low emotional awareness in BPD should focus on bodily signals and could very well involve biofeedback [6,12-23]. Although there are indications that emotional awareness can improve with psychological interventions [24], development and testing these interventions is in its infancy [25,26]. From literature and information gathered from patients and professionals, we identified a need to improve the treatment of low emotional awareness in BPD [19,24,27].

We started a project to develop a biosensor-informed e-coaching app on emotional awareness in the challenging environment of a psychiatric ward for patients with severe BPD. We gave the app the name *Sense-IT*, as it refers to both its intended purpose for patients (*sense it*) and its *technological nature* (IT: information technology). Importantly, it should help its users to learn to better recognize changes in their physiological and, with that, emotional arousal and thus increase emotional awareness. To the best of our knowledge, this project is one of the first in providing ambulatory biofeedback to this group of users. We previously developed an initial prototype [28]. The aim of this study was to complete the next step and deliver a working version of the Sense-IT app that is deemed useful and usable by 3 important groups of stakeholders: patients with BPD and low emotional awareness, mental health professionals working with these patients, and experts on user-centered design (UCD).

Initial Prototype

We decided to design our app for wearable technology that comes equipped with essential biosensor technology that is widely available for consumers, is affordable, and runs on a mature operating system (OS) that offers easy-to-use app programming interfaces (APIs) by which one can develop native apps and access sensors directly. After a small preliminary study with main stakeholders and users on acceptable, usable, and nonstigmatizing hardware (unpublished), we decided to use a smartwatch and mobile phone. Next, we decided on developing an app for Google's Android (including Wear) OS. As a mature, easy-to-access OS, it has a broad base of support by manufacturers, software developers, and users alike. Most smartwatches (at the time) came equipped with a photoplethysmogram (PPG) sensor and accelerometer. The PPG sensor is used for measuring heart rate (HR). After studying the literature on the use of biosignals for affect detection, we built an algorithm that takes the HR data of the user to calculate a *physiological correlate of emotional arousal* (*PCEA*). HR, similar to HR variability (HRV) and electrodermal activity (EDA), is triggered by the sympathetic nervous system [29,30]. Typical consumer wrist-based HR monitors provide a fairly accurate measurement of HR even when deployed during physical activity and movement [31].

The first version of the app was built by one of the authors (RK) and 2 graduate students in Human Media Interaction. The outcomes of previous design cycles served as the main sources of input for programming (refer to the study by Derks et al [28]). Visual design was kept low in complexity so that the app's graphical user interface (GUI) is easy to interpret by the user yet remains nondescript to the environment (see Figure 1). To cater to the expressed need for discreet and unobtrusive coaching, audio signals were deliberately not implemented. Instead, the app was programmed to give tactile feedback via the vibration motor in the smartwatch. Users have direct access to a proxy of the current level of the PCEA via continuous HR measurements on both the smartwatch and mobile phone. On the smartwatch, users have the option to choose from 4 different watch faces. The mobile phone app provides an overview of the recent measurements. At higher levels of arousal (stages 4 and 5 out of 5), the user receives a textual prompt, intended to stimulate deliberate reflection on one's current status of emotional arousal. See Figure 2 for an impression.

An option to manually export recorded data to a server was also implemented. With this option, a primary user is able to transfer his or her data to a server. These data can then be accessed via an internet browser through a secured gateway. The data can be presented in raw form or plotted to provide a graph depicting the changes in PCEA over time. This option was built in to increase the number of ways the data could be used to support therapy. For example, it could provide an opportunity for the patients to get a better overview of the changes in their arousal over time or to let their therapists have a view (if they were granted access to the secured data by the patient). Although this functionality was built in, it could not be tested at the clinic because of limitations in the technical infrastructure at site and ethical limitations concerning the sending of personal data to servers outside the clinic. Figure 3 gives an overview of the setup of the app.

Figure 1. Sense-IT mobile phone graphical user interface, first version. The yellow switch starts and stops the measuring of heart rate (HR) via the smartwatch. Each recorded change in level of HR (or physiological correlate of emotional arousal [PCEA]) is represented a blue bar in the white area of the screen. The date and time of the change are displayed within the blue bar, and the level of the PCEA appears as a number in the circle left of the bar.

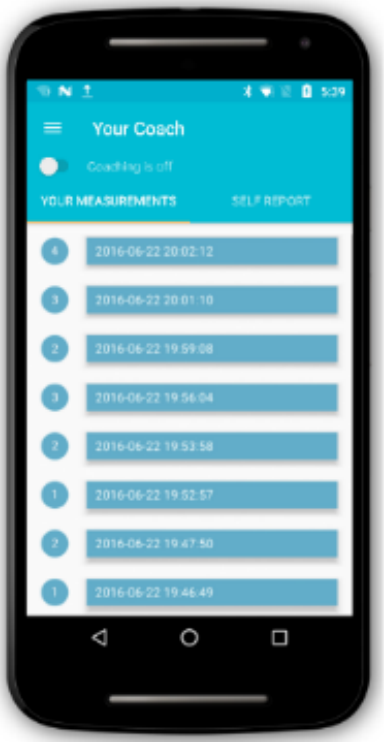
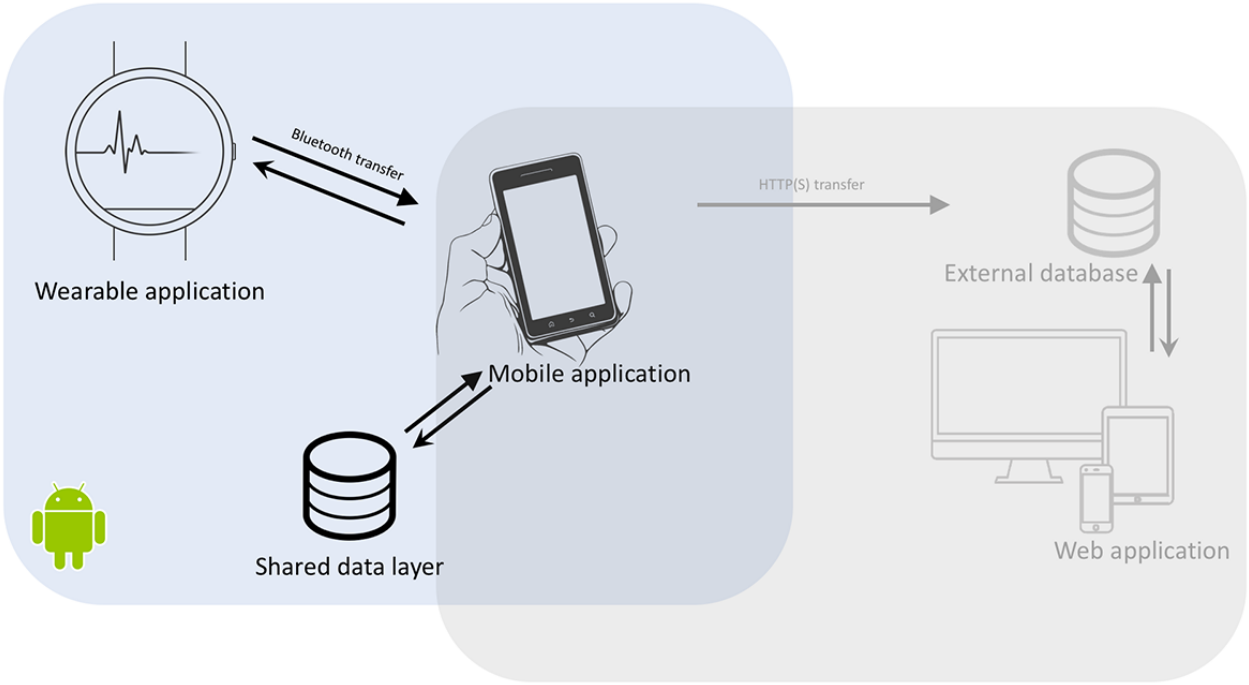


Figure 2. Sense-IT watch faces, first version. The watch faces each expressed changing levels of HR/PCEA by changing either the circular elements (ie, more circles or a bigger circle), or simply indicating a numerical level (from 1 – 10).

Figure 3. Overview of the setup of the Sense-IT app.



Methods

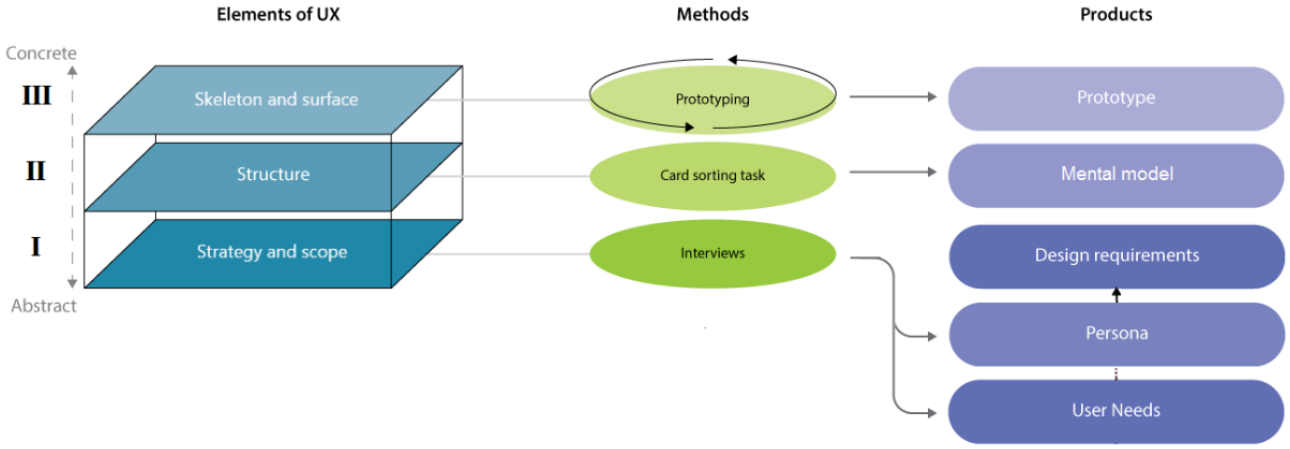
Study Design

The Methodological Framework

A design framework tailored to the setting and purposes was developed based on principles from user experience (UX) design and UCD. We previously published a detailed report on the development of this framework and our first steps in the development of the Sense-IT app [28]. We combined the conceptual overview of *Elements of UX model* [32] with the broader focused approach of the *CeHRes Roadmap* [33] and added a placeholder to specify the methods used. This led to the EMP framework. *EMP* stands for *element-method-product*, the main components of the model. Figure 4 shows a schematic overview of the EMP framework.

We previously completed the steps of the first 2 elements of the framework, that is, the *strategy and scope* and the *structure* plane (in Figure 1, I and II of the elements of UX) [28]. The completion of these steps in the model yielded design requirements and mental models that serve as the basic input for the current, third element of the framework (in Figure 1, III, *skeleton and surface* of the elements of UX). In the third element, the desired product was a working prototype. This implied getting into the back-end organizational structure and flow (*skeleton*) and the front end or outer qualities such as visuals, sounds, and vibrations (*surface*). As the main UX method (the *M* in the EMP framework), we chose cyclic, iterative prototyping [34].

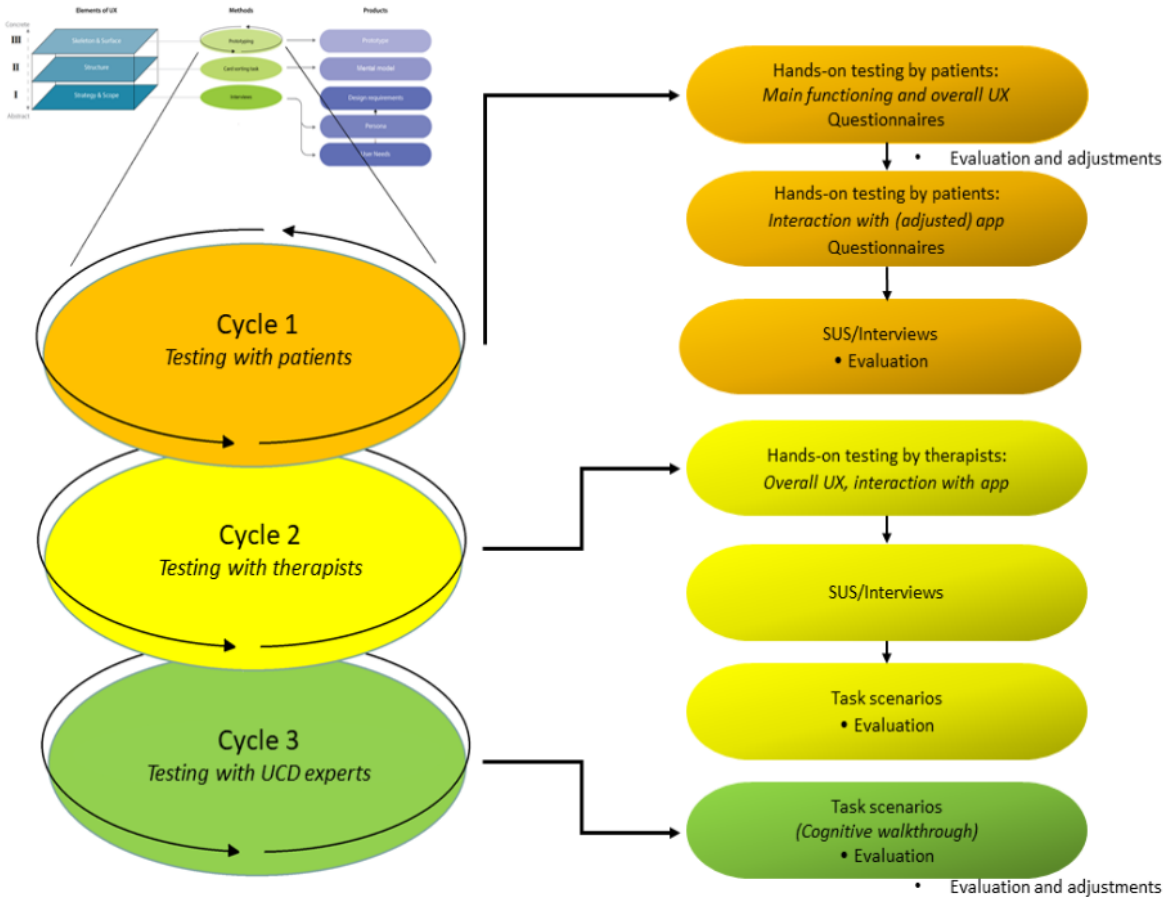
Figure 4. The EMP-framework. The framework is to be read from the bottom (ie, abstract considerations) to the top (ie, concrete considerations). The lines indicate the connection between the elements and the methods. The arrows point to the products that will result from the methods.



Procedure

We went through 3 cycles of testing, each with its own specific group of users (see Figure 5). The first cycle, that is, initial prototype programming and pilot testing with patients, took place between February and July 2016. The second cycle, that is, usability testing with professional caregivers, took place between November and December 2016. The third cycle, that is, usability testing with expert users, took place between January and February 2017.

Figure 5. Specification of the Method of Prototyping within the EMP-framework.



Setting

The first 2 cycles took place within an inpatient psychotherapeutic setting of Scelta, GGNet, a mental health care provider situated in the east of the Netherlands. Testing during the third cycle took place in an office setting within the facilities of the University of Twente.

Scelta is a specialized division within GGNet, one of the larger mental health care providers in the Netherlands. Scelta provides psychiatric treatment of severe personality disorders. At their inpatient clinic, Scelta offers several psychotherapeutic treatment programs for personality disorders. The testing of the app with patients and therapists was within the dialectical behavioral treatment (DBT) [35] unit of the clinic. Here, multidisciplinary group DBT [35] is given to a maximum of 27 inpatients at a time. All patients are diagnosed with 1 or more personality disorders—the majority of them having BPD diagnosed as main disorder. Four days a week, during working hours, patients receive multiple forms of psychological treatment. Medical treatment can be part of the treatment but serves a subsidiary role. On average, duration of treatment within the DBT program is 9 to 12 months. After that, most patients are referred for further (outpatient) care.

Data Analysis

Data from interviews, questionnaires, and task scenarios were transcribed via a transcription program (*F4*). Additional remarks made by participants during or in between hands-on testing were also registered. Qualitative analysis consisted of a combination of content and thematic analysis [36]. Comments were categorized by applying constant comparison, comparing statements numerically and content wise. Usability problems and needs were listed and subsequently grouped, after which a central theme was allocated to each group. This was done by 2 of the authors (YD and RK), of which one has a background in psychology and the other in computer science. Final categorization reflects consensus after comparison and discussion of the interpretation processes of both authors. The following main themes were identified: (1) technology, (2) user interface and interaction, and (3) functionality (and ethics). Next, within each set of results per group, inductive thematic analysis was used to identify whether, and, if so, which specific values regarding the acceptance and adoption of the app were present. Next to the qualitative analyses, individual and overall scores of the system usability scale (SUS) [37] (used in cycles 1 and 2) and the cognitive walkthrough (used in cycle 3) were quantitatively analyzed.

Ethical Considerations

The study was granted approval by both a nationwide operating ethical commission and a local ethical commission associated with the university. Participation in the study was on a voluntary basis and only after informed consent. Patients undergoing clinical treatment at the treatment center and therapists working in the center were eligible if they were willing to actively participate. All participants could withdraw from the study at any time with no further obligations. In addition, therapists of the patients held the right to anytime exclude or withdraw patients from participating in the study if they judged a patient's participation potentially detrimental for his or her well-being or unwanted or inappropriate in any other way. There was no financial reward or other incentive for participating.

Cycles

As the cycles were carried out as three consecutive studies, each with its own participants, procedure, and materials, the remainder of the method section will be discussed per cycle. Next, results will also be presented per cycle, in paragraphs representing the main themes that resulted from the thematic analysis. Multimedia Appendix 1 pprovides a schematic overview of the thematic ordering. It lists all usability issues and user needs that were identified by 1, 2, or all user groups grouped by the main theme and presents the applied—or planned—solutions and adjustments to the Sense-IT app after the 3 cycles of usability testing. The result section concludes with the presentation of the current version of the Sense-IT app.

First Cycle: Testing with Patients with Borderline Personality Disorder

Participants

A total of 5 patients of the DBT program participated in testing the app. Participation was on a voluntary basis and after informed consent. All were previously diagnosed with BPD and low emotional awareness. Of 5 patients, 4 completed the second iteration, and 1 patient could not participate on the second day of the second iteration, as she forgot that she had to attend other meetings. Patients were aged between 18 and 49 years (mean age 28 years, SD 11.82). All participants were females and all had Dutch as their native language.

Procedure

The first cycle consisted of 2 iterations, each with a design phase and an evaluation phase. During the hours of testing, members of the research team remained standby at the clinic in case a patient would encounter technical issues or would come up with questions regarding the app.

During the first iteration, the focus was on the main functioning and overall UX of the app. Patients received a short explanation of the system and some information on the upcoming days. On the first day, patients just had to wear the smartwatch to gather the required HR data to set personal baseline values. They did not have to interact with the system but were asked to monitor their personal experiences on wearing *the hardware*. The second day, patients again wore the equipment running the Sense-IT app. They were asked to use and interact with the Sense-IT app as if it was an actual adjunctive to their therapy. Patients were asked to fill in a questionnaire at the end. After the first iteration, adjustments were made to the app.

During the second iteration, the focus was on patients' preferences on the (graphical) user interface ([G]UI) and the interaction with the system. It again consisted of 2 days of testing. Patients were instructed to interact with the app as if it was part of their therapy program. The main goal of the second iteration was to examine whether the alterations to the app led to a satisfactory overall UX. At the end of the day, patients were asked to fill in the SUS. They also were briefly interviewed on their experiences.

Materials

Hardware

A total of 5 mobile phones and 5 smartwatches were used in sets. Each set consisted of a Moto G, third-generation smartphone (5-inch screen size and screen resolution 1280×720 pixels) and a Moto 360, second-generation smartwatch (1.37-inch screen size and screen resolution 360×325 pixels). Both ran a version of Android OS: Android 6.x (KitKat) on the mobile phone and Android Wear 1.5 on the smartwatch. Each set was connected via Bluetooth (and optionally via Wi-Fi) through built-in communication software provided by the Android ecosystem.

System Usability Scale

The SUS is a commonly used questionnaire that quickly and reliably assesses the usability of a product [37-40]. The SUS [39,40] comprised 10 statements that are scored on a 5-point scale, ranging from *totally disagree* to *totally agree*. It contains statements such as "I thought the system was easy to use" and "I thought there was too much inconsistency." The SUS yields an overall score of the system usability ranging between 0 and 100, where higher scores indicate better usability. For interpretation of scores, we used the guideline provided by Bangor et al [38].

Questionnaires

A total of 2 short self-constructed questionnaires were administered to the patients during the first cycle of testing. The first questionnaire (26 items) contained questions on the use context; the patient's level of experience with technology; general UX of the app; UX of the interface on mobile phone and smartwatch; the use of prompts; and additional questions regarding privacy, perceived risks, and missing/desired features. See Multimedia Appendix 2. The second questionnaire (16 items) contained questions on the patient's general UX with the mobile phone and smartwatch interface, the use of the diary/note keeping function, and questions about the option to share data with therapists in the future. See Multimedia Appendix 3.

Interview

After each iteration, a semistructured interview was held with each patient. The interview consisted of 5 open questions. Participants were asked what they liked about the intervention (app plus hardware), what they disliked about it, what their experiences were regarding the measurement of their bodily signals (PCEA), if they would have made other decisions regarding the design of the app and/or choice of the hardware,

and if they had any other remarks about what they would like to see improved or altered in a future version.

Second Cycle: Testing with Therapists

Participants

Of 25 health care professionals at the clinic, 4 were invited to participate. They were selected via a stratified sampling method [41]. The sample included 1 psychiatrist, 1 bodily oriented psychotherapist, and 2 groupworkers or sociotherapists. All were native Dutch and aged between 47 and 62 years (mean age 52 years, SD 7.1). Overall, 2 participants were males and 2 were females. One of the participants personally owned a smartwatch. In addition, 3 of the participants indicated to have affinity with wearable technology but indicated not to closely monitor developments in the field of smart devices.

Procedure

The app and the 2-day testing procedure were kept similar so that the health care professionals would get a similar UX as the patients. On the second day of testing, a series of 4 paper task scenarios were completed along with the SUS and a short interview.

Materials

Hardware

The same mobile phones and smartwatches were used as in cycle 1.

System Usability Scale

See the description under cycle 1.

Task Scenarios

A total of 4 scenarios were written by the research team, based on the functions of the app. Each scenario consisted of a task to be performed on either the mobile phone or smartwatch: (1) *add a comment to your latest measurement*, (2) *change the watch face to another Sense-IT watch face*, (3) *add a general comment to the timeline (not to a measurement)*, and (4) *let the app stop measuring your HR*. With each scenario, the user was asked which steps he or she took to perform the task, what difficulties were encountered while performing the task, and whether the user had further comments or suggestions.

Interview

To close off the 2-day hands-on testing phase, a brief, semistructured interview was held with each therapist. The interview questions were the same as in cycle 1, but the interview also inquired how the app could be implemented in the therapies of their patients.

Third Cycle: Testing with User-Centered Design Experts

Participants

For the third cycle, 3 expert users were personally contacted. All were part of the professional network of the authors at the University of Twente. Of 3 expert users, 1 works as an assistant professor in Human Centred Embodied Design, with expertise in the field of assistive technologies. The second expert user is a researcher with a background in Biomedical Engineering who currently conducts research on telemonitoring in a medical setting. The third is an assistant professor in Product Interaction Design whose current work is focused on multisensory design, user interaction and experience, and their influence on behavior and motivation.

Procedure

The UCD experts were asked to complete a cognitive walkthrough while using the app in an individual session. They were provided with information on the primary users and their concerns in the form of a persona [28]. After the cognitive walkthrough, each UCD expert was given a final moment to reflect and mention any other detected usability problem they had encountered.

Materials

Hardware

The same mobile phones and smartwatches were used as in cycles 1 and 2.

Cognitive Walkthrough

The UCD experts were asked to complete several tasks while thinking out loud about what a primary user (ie, a patient with BPD and low emotional awareness) would do and evaluate if the task at hand would be easily achievable. The tasks were the same as the scenarios for the therapists.

There were 4 additional questions on usability to be answered with *yes* or *no*. These questions were as follows: (1) *Will the primary user try to achieve the correct effect?*, (2) *Will the primary user notice that the correct action is available?*, (3) *Will the primary user associate the correct action with the desired effect?*, and, if the user performed the right action, (4) *Will the primary user notice that progress is being made toward accomplishment of her goal?* As an indication of usability error [41], the average number of *yes* answers were added up to a score between 0 and 1, with higher scores indicating less errors.

Results

First Cycle Results

Quantitative Data (System Usability Scale)

The patients rated the system after the second iteration of the first cycle. The overall rating was *good* (average score of 78.8). Of 4 participants, 3 rated the app as good to *excellent* (85 to 97.5), whereas one rated it as *poor* (42.5) [38]. The latter participant experienced trouble in working with the mobile phone and smartwatch in general and indicated that her age could play a role. She would therefore have liked to see the user interface simplified and more *foolproof*. She also would have liked to have an easy to understand user manual to come with the app.

Qualitative Data (Questionnaires and Interview)

Technology

Overall, patients were positive about the ease of use of the app on the smartwatch and mobile phone. They also liked the design

and indicated they had no problems with having to wear a smartwatch—although not all of them were used to wearing a watch. Moreover, 3 patients found it cumbersome that the app was on a mobile phone of the project, so they had to carry it besides their own device. Given the test phase, all understood why the app was not installed on their own devices.

User Interface and Interaction

All patients experienced the app as discreet and inconspicuous. All would have liked more data visualization. They stressed it was important that the data visualization should remain *neutral* for reasons of discreetness while still being intuitively understandable. Regarding the Watch face GUI, users found it important that they could choose from several designs.

Patients stressed that the app should be compatible with their therapies. This need was already addressed, as physiological arousal was represented in a *two times five* stage process that is compatible with the current user setting (DBT and systems training for emotional predictability and problem solving [3,42]) as well as cognitive behavior therapy (CBT). Patients indicated that other ways of breaking down in levels would still be welcome for the use of the app in other forms of therapy. One of the patients believed that the levels of the physiological scale were representing the subjective emotional states that were used in the main model on emotions in therapy.

The timing and frequency of feedback given by the app (via the vibration motor in the smartwatch) yielded mixed reactions. One patient said the notifications matched the perceived moments of arousal well. For 2 patients, the frequency of notifications was acceptable. One patient received *a bit too many notifications*, especially since they were accompanied by a vibration of the smartwatch that was noticed by other patients during group therapy. In contrast, another patient received too little notifications in regard to her experienced moments of heightened emotional arousal. She also believed the notifications were sometimes lagging behind with her actual feeling of increases in arousal. The reliability of the app was perceived as average to slightly unreliable. One patient encountered a problem in HR measurement, which was discovered and adjusted the second day.

Functionality (and Ethics)

One of the main findings after the first iteration with patients was the expressed need for an option to add a note to a registered change in PCEA level by the app. Regarding privacy, all patients answered that future sharing of the information generated by the app with their therapists would not be an issue for them.

Adjustments and New Features Added After the First Iteration

To prevent confusion of *actual*, subjective emotional arousal and the PCEA provided by the Sense-IT app, the numerical denoting of the scales was changed to icons consisting of spheres. Here, 1 sphere is equivalent to level 1, and 5 spheres are equivalent to level 5. One watch face was replaced by a new one, as the original one was judged as little attractive.

Next to adjustments in graphical layout, the app was extended with 4 new features. Of 4 features, 2 were based on the feedback received from the primary users, and the other 2 were based on the review of the app by the research team. The first new feature was the option for the user to provide feedback to the system by adding personal notes to each notification of change in PCEA by the Sense-IT app.

The second new feature was the option to add personal notes to a *diary* that is unrelated to recorded changes in PCEA level. As this is a common element in regular psychological treatments such as CBT, the idea for adding a diary was to enable patients to comment on their experiences over longer periods of time, for example, their morning, whole day, week, instead of adding comments to momentary situations captured in the app. This is also in line with recommendations from a recent review study of mobile health (mHealth) mobile phone apps [43]. Both features were added in response to an expressed need for a memory aid. Such an aid could be used, for example, during appointments with professional caregivers.

The third feature was a settings tab in which several personal values/preferences could be set. The addition of this feature was based on a review of the app by the research team. It concerned the option to manually adjust the mean HR and SD used by the system, as well as an option to alter the *sensitivity* of the system, that is, decrease or increase the threshold of when a *new level* was reached. We implemented 3 levels of sensitivity: *normal* (every change in HR of 1 SD adds or subtracts a PCEA level), *low* (change of 1.5 SD), and *high* (change of 0.5 SD). The option to set the sensitivity of the app was implemented to better adjust to the preferred level of feedback received by the user. The option to manually alter the values of mean HR and SD was primarily added to let the researchers override the values set by the system based on the baseline measurement in case the set values were unrepresentative for the user (eg, the user turned out to have an uncommonly low or high average HR during baseline measurement). The perceived *lagging* of feedback that 1 primary user reported could not be addressed in this stage of development but was scheduled as future work.

A fourth new feature was the ability of the app to also measure and represent PCEA states up to (theoretically) 5 steps below the user's average HR. The decision to add this feature was made by the research team and was based on current expert theory on emotional (under)arousal in the field of affective neuroscience and psychotherapy [44]. These states were visualized in the app as hollow and/or blue-colored spheres (see Figure 6).

Figure 6. Examples illustrating the visualization of the low-arousal state "hollow spheres" on the smartphone (A) and smartwatch (B) interface.

Second Iteration

The adjusted version of the app was used in the second iteration. All patients indicated that they experienced the option to add notes to detected changes in PCEA as useful. Of 5 patients, 3 had used the feature on more than 1 occasion. The diary for miscellaneous notes was also perceived as useful by all patients, although none of them had actually used this function.

The representation of PCEA states with below the personal baseline was also regarded as useful: "By adding this feature, the therapeutic goal to reach or maintain an overall lower state of arousal is supported by the app." Patients suggested that they could practice mindfulness or do an exercise from their relaxation training and see if it had a lowering effect on their level of arousal.

All patients indicated that the use of notifications increased their self-awareness. Although 2 of them indicated that the app initially induced some more stress, for example, by making them become too focused on the feedback from the app and reacting to that, all indicated they quickly got used to *being monitored continuously*. All indicated they would like to see this app being integrated into their therapy, preferably as an adjunct that could be used in regular face-to-face meetings with their therapists. They all thought the app had positively affected their awareness of their emotions. One patient indicated she got notifications about increased physiological arousal during group therapy, which led her to observe her emotional arousal more closely. She thought this had really helped her to make more out of the session. All participants answered they would have liked to continue wearing the devices and using the app all day.

Second Cycle Results

Quantitative Data (System Usability Scale)

The therapists rated the system with SUS scores between 30 and 85, with a mean of 59.4, which is *OK* for the usability of the app, but *a candidate for increased scrutiny and continued improvement* for passable products [38].

All therapists described the app as useful. The lowest SUS score was because of a reliability issue, as she did not receive any notification during testing. However, she still expressed a positive attitude toward the app: "The app did not function properly, so that's a main reason why I reported negative experiences [on the SUS; authors]. However, I still welcome the therapeutic function the app could fulfill, so keep up the good work!"

Qualitative Results (Task Scenarios and Interview)

Overall, the feedback by the therapists conveyed as a general message that the app bears real potential but should first be disposed of all bugs and errors. They all stressed how important it was the app should not attract *unwanted attention* in any way when integrated in daily practice, either meaning it should not hinder their daily work as a therapist (eg, having to do *extra work* when the app should start malfunctioning) nor disrupt standard forms of therapy (eg, a notification by the app draws the patient's attention, which disrupts the process of face-to-face therapy) or draw unwanted attention to the patient in social settings.

Technology

The therapists mentioned that every now and then the system seemed to stop processing HR data. Overall, 2 said such technical difficulties were experienced as highly demotivating

and could cause stress with patients. In addition, 3 therapists mentioned that it was unclear to them how far apart the smartwatch and mobile phone can be without losing the Bluetooth connection, which caused a feeling of uncertainty. One suggested to add a manual (in the app or just on paper) that provides such information. One also suggested that it would be better to have multiple sources of physiological data, as this could improve the accuracy of the PCEA.

User Interface and Interaction

The general layout of various buttons proved less then optimally intuitive for 2 of the 4 therapists. During the scenarios, they found the option to add a miscellaneous note confusing when asked to add a comment to the latest registered change in PCEA. All indicated that the *self-report* tab was easy to find and leaving a note was easy to do. Also, ending the measurement of HR was not clear to the 2 previously mentioned therapists. Although the other 2 immediately correctly pressed the yellow switch to stop the measurement, 1 of the other 2 did not manage to find the on/off switch at all.

Furthermore, 2 of the therapists found the graphical representation in spheres convenient and logical. The other 2 therapists, however, did not figure out the meaning of the spheres by themselves. In addition, for both these therapists, it was not clear what the buzzing of the device intended to convey. They did understand how the app worked after it was explained to them by the researchers. One of the 2 therapists who understood the app *by intuition* commented that the app should be even more *graphical*: it should present the user with more graphs and figures.

Functionality (and Ethics)

All therapists mentioned it was of significant clinical relevance that the app communicated not only states of PCEA that were higher than the primary user's baseline but also those that were below the personal mean. This way, patients can comment on their *physiologically calm* moments as well and learn from it. None of the professionals was able to change the interface of the smartwatch. The set of operations required by the user proved to be nonintuitive and prone to errors, resulting in the app to stop working.

Third Cycle Results

Overall, UCD experts considered it to be easy for users to access information and perform actions but also gave several suggestions for improvement of the app.

Quantitative Data (Cognitive Walkthrough Evaluation)

The average scores per task on the cognitive walkthrough ranged between 0.83 and 1.00 and the average score across tasks was 0.87. These scores indicate that the app's overall usability is somewhere between "OK" and "acceptable," yet could be improved.

Qualitative Data (Cognitive Walkthrough Evaluation)

Technology

A problem indicated by all 3 experts was that the collecting and processing of sensor data by the app seemed to stop at random moments, although they were not certain if this was really the case. When the Sense-IT app is turned off within the mobile phone environment, the user sees a message on the smartwatch that says the app is waiting to receive data from the sensors. This was judged as confusing to the user, as it does not specify whether the app is idle and awaiting an action by the user or it is waiting for sensor data that will be transferred automatically.

What was found missing was an electronic or paper manual. One of the UCD experts indicated that not all functions/features of the app are intuitively clear to the user. He suggested to include a manual with information on how the app works, what the different number of circles mean, and what a user can do with this information.

User Interface and Interaction

Navigation

Overall, the experts judged the interface to be low in complexity and easy to comprehend. Working with the app was judged as simple, and most tasks were easy to perform. Still, navigation within the app could be improved. Several small flaws negatively influenced the flow and experience of the interaction. Changing the interface of the app on the smartwatch was judged as intuitive, however, *only* when users are familiar with using a smartwatch or mobile phone. The option to go back to the previous screen was sometimes hard to find. The location of the on/off switch could be better. As people typically search for an on/off switch on the upper (right) corner of a device (such as a remote or a mobile phone), it was advised to place it there in the app as well. It was considered an additional issue that the smartwatch did not give visual or haptic feedback to the user when the collection of sensor data stopped.

All 3 UCD experts considered it easy to figure out how to add a comment to the latest registered change in PCEA. For added clarity, the text *latest measurement* could be added on top of the column. One expert suggested to link the *self-report* function on the main screen of the app to a button, not a separate tab. All expressed doubts about the usefulness of the option to add miscellaneous notes apart from adding notes to detected changes in PCEA. It was suggested that the functionality and UX could be improved by integrating both in 1 timeline.

Visual Layout

Another navigational issue concerns the visual layout of the homepage. When the user is on 1 page, the other page/tab is displayed in light gray shading. This might suggest that this function is not available. On the opposite, the color of the text *coaching is on/off* is the same as the active buttons, which suggests it can be switched on and off by touching it, while it cannot. The experts suggested that a change in colors or layout can solve this inconsistency.

Textual Layout

Next, some windows or buttons contained words that were too *psychological* or *technical* and thus were difficult to understand for most users. Examples were the line of text "You are high in your physiological scale" and the term *self-report*.

Personalization

The UCD experts made several suggestions to enhance the level of personalization within the app. In the settings menu, a user identification is displayed in the current version of the app with a number that does not mean anything to the user. It was suggested to replace this number with the option for users to add their own name. Other suggestions were adding options to switch contrast of the GUI (black/white conversion) and/or the option for the user to personally set the colors or choose between several color schemes.

Persuasiveness

The experts gave several suggestions to increase persuasiveness. A suggestion to persuade the user to add notes and comments was to include changing colors (a detected change with a note gets another color than one without) or by displaying a commenting space with each detected change in PCEA, so it is directly visible where personal notes have (not yet) been added. It was also suggested to include on-screen notifications on the mobile phone and visual reminders on the smartwatch. When receiving an on-screen notification, a comment could then be added without having to open the actual app. The experts suggested to have the app actively request to add a comment at more levels of PCEA. In the tested version, the app only does so when the user's PCEA reaches the fourth or fifth state. One expert advised to further integrate the use of emoticons. At the moment, the app already allows the user to add emoticons to a note, but a suggestion was to add a separate column for emoticons in the notes section.

Visualization of Measurements

Several suggestions for improvement were about presenting the user a more *graphical* overview. In the tested version of the app, the overview of all measurements is fairly textual. To improve the overview for the user, colors could be added such as, for example, darker blue on the higher emotions. Also, a graphical timeline with a scalable timeframe (day, week, and month) could be implemented to track the PCEA.

Functionality (and Ethics)

A suggestion was to add the option to take pictures or record videos via the camera of the mobile phone. In the settings menu, not all options were clear to the UCD experts.

Sense-IT: Current App

On the basis of the analysis of the identified usability problems and user needs (and proposed solutions) gathered after completing all 3 cycles, the Sense-IT app was revised. The app as a whole still consists of a wearable app on a smartwatch and a mobile app on a mobile phone, all implemented using the Android ecosystem. Below, we describe its components and functionalities.

The Mobile Phone App

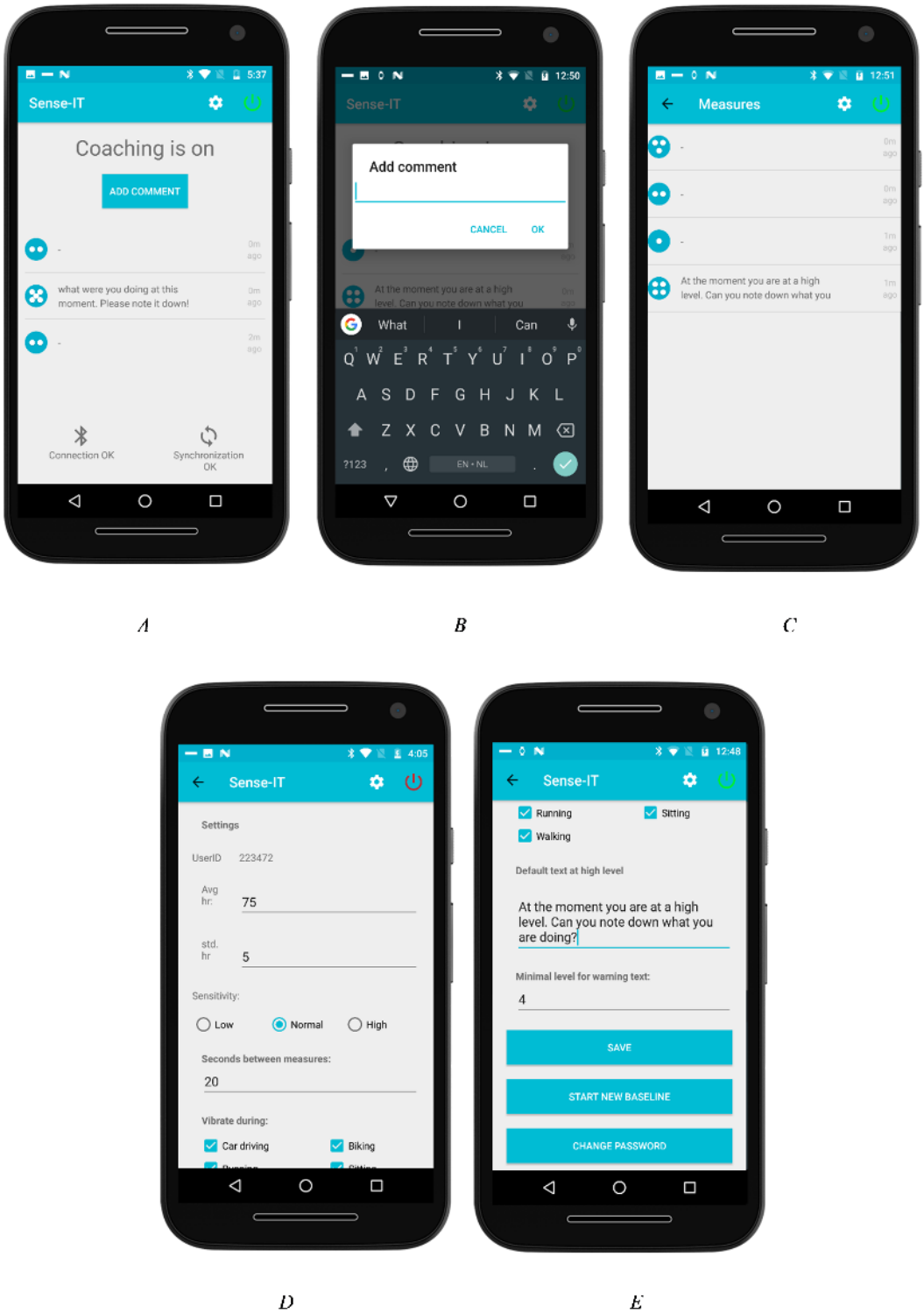
The mobile phone app of the Sense-IT system consists of 3 components: (1) communication and storing of data from the wearable app, (2) the algorithm detecting changes in physiological arousal, and (3) the user interface.

The *first* component reads the data pushed from the wearable app and stores the data in a local database on the mobile phone itself. The algorithm then evaluates changes of PCEA based on new available data. By default, new and old data are compared every 10 seconds. The time between evaluations can be altered (limited by the specifications of the hardware).

The *second* component is the algorithm and takes into account the personal average HR and SD of the user and the current activity and the current HR of the user. It classifies the PCEA to 1 of 10 levels ranging from −5 to −1 and from 1 to 5. The app notifies users when their HR (measured via PPG) decreases or increases markedly (the boundaries are determined by the user's mean HR and SD, but these values can be personalized). The average HR and SD are based on the results of a baseline measurement in which HR is measured until a preset number of valid measurements (standard setting is 300 measurements) is collected. Notified changes will be mostly unrelated to physical activity, as the app will only notify the user when he or she is not involved in vigorous action (as determined by the onboard accelerometer and associated activity recognition algorithms). All changes in PCEA, together with the classified type of user activity, are recorded and displayed in the overview of the mobile phone app.

The *third* component, the user interface, supports interaction between user and app. The dashboard page of the app (Figure 7A) presents an overview of the status of the app, for example, the status of the connections and synchronization. The last 3 detected changes in arousal are displayed, and there is an option to add notes. Clicking it brings up a new window where notes can be added (Figure 7B). Users can turn on the app by pressing the on/off icon in the top-right corner of the user interface. By clicking on *show more* (which appears in a box below the last 3 detected changes once there are more than 3 recorded changes), the user opens a timeline of all changes detected by the system (Figure 7C). The events are listed in chronological order. The events in the timeline are displayed with their level, the message written by the user, and the time when the event did happen. By clicking on one of the events, the user can add a note or edit a note that was stored (see Figure 7B). The (scrollable) settings page of the app is opened by clicking the settings icon on the dashboard (Figure 7D). To prevent changes in the settings by unauthorized users, this page is password protected. Within the settings menu, the user can (re)start a new baseline measurement (see Figure 7E). Measured values can also be manually adjusted if needed. Other adjustable settings include the sensitivity of the algorithm (high-medium-low), the time by which the algorithm checks for changes in physiological arousal, and the option to select the type of activities whereby the app should or should not give a notification when a change of physiological arousal is detected (see Figure 7D). Users can define their own message that will be displayed when their PCEA reaches a predefined level. This level can also be set by users in the setting page (Figure 7E, both options are in the middle section of the screen).

Figure 7. UI of the current version of the Sense-IT smartphone app (A-E).



The Wearable App

The wearable app uses the accelerometer (movement) and PPG (HR) sensors on the smartwatch to monitor the user. Sensor events are registered as fast as possible by the hardware via the Android sensor manager API and registered on the smartwatch app. With the hardware used in this study, this comes down to once per second. A sensor event contains data fields associated with the event. In the case of an HR registration, these data fields include the sensor that generated the event, the accuracy of the event, the timestamp of the event, and the value of the event in beats per minute. The accuracy value is a value between −1 (no contact) and 3 (most accurate) that is determined by Wear OS. The Sense-IT app is currently set to include all measurements with value 1 and higher. These measurements are sent through the device APIs as data events and are received by the mobile phone for further processing.

Figure 8. Current watch faces.

The wearable app is also responsible for presenting the GUI on the smartwatch. On the basis of data from the sensors and the result of the algorithm of the app on the mobile phone, the screen of the watch will present the current state of the measurements via 1 of the 4 available watch faces. Figure 8 shows the current watch faces. The user can alternate between them and choose the one they like via the Wear OS settings.

Discussion

Primary Findings

This paper describes the development and usability testing of Sense-IT, a wearable biofeedback app for Android-based devices that can be used in the daily practice of a mental health clinic for personality disorders. The app is meant to support BPD patients in increasing their level of emotional awareness. The app was tested on usability by patients, therapists, and UCD experts.

The concept of a biosensor-informed app for emotional awareness enhancement was appreciated by both patients and therapists; the prototype was judged as promising by all user groups. After 1 iteration in the first cycle of testing with patients, basic functionality of the Sense-IT system was rated as acceptable. This waived the need for a major revision before starting the consecutive rounds of testing with mental health care professionals and UCD experts. Still, results after finishing all 3 testing cycles made clear that the app should be considered a *candidate for increased scrutiny and continued improvement* [38]. In total, 30 usability problems and/or needs were identified (Multimedia Appendix 1). All 3 user groups brought up usability problems and suggestions for improvement. UCD experts identified most of these usability problems (20 in total). Patients identified 11 issues, and therapists identified 9 issues. There was some overlap between them: 8 issues were mentioned by at least 2 groups. The UCD experts brought to our attention 14 themes or problems that were not mentioned by the other groups, the patients 6, and the therapists 2. Most usability problems and needs could be addressed in the software revision that followed after the 3 cycles were finished. This resulted in the version of the Sense-IT app presented at the end of the results section.

Results of the 3 cycles of testing favor the use of ambulatory biofeedback to improve emotional awareness in patients suffering from BPD and low emotional awareness. To our knowledge, the Sense-IT is one of the first scientifically grounded apps that can be used in clinical research and/or clinical settings for longer periods of time without requiring extensive support by researchers and/or developers. Although it may be a quite simple app from a technological perspective, from a mental health perspective, it is a real innovation. This applies to both the way it was developed as to it being a new way of delivering treatment to patients. The use of consumer

technology to enable *always available* indices of physiological changes could prove to be a relevant addition to existing therapeutic interventions, even if the measurements and algorithms used are relatively crude and simple, and the integrated sensors have limited resolution and less than perfect accuracy. There is an ever-growing number of publications that introduce and discuss concepts of biosensor-informed mHealth interventions on emotional/psychological awareness [45-47]. However, at present, *feeling the changes in the body* is something patients with mental health problems still mainly have to learn by purely subjective methods. We found 1 recent project in which—nonwearable—technology was used in a mental health setting that aimed to "direct alexithymic persons to reflect on their internal, somatic experiences as a source of information for interpreting and labeling emotional experiences" [15]. Although there are numerous commercial companies and startups that sell products and/or apps that claim to help to raise awareness on—or even directly measure—emotion, emotional arousal, or stress, they generally lack research that validates or supports their relevance for users and/or validity (see the study by Peake et al [48] for just a select number of examples).

Limitations and Strengths

Of course, both the app and the study have their limitations. Regarding the choice of hardware, research-oriented hardware that gives access to more advanced or potentially *better* sources of physiological data such as EDA and/or HRV exists [49]. However, the use of HR as the main cardiovascular parameter for physiological arousal is a defendable option. HR, such as HRV and EDA, is triggered by the sympathetic nervous system [29,30]. Typical consumer wrist-based HR monitors provide a fairly accurate measurement of HR even when deployed during physical activity and movement [31]. Measuring HRV is much more susceptible for producing artifacts under real-world conditions. Even producers of wearable technology such as the E4, which is claimed to be able to measure HRV from PPG, stress that this is only feasible in short scenarios (ie, several minutes) that are free of movement.

We believe that with time, better (consumer) hardware with more and more advanced sensors will become available, as will be the case for more accurate signal processing algorithms that can be used in (validated) wearable devices [50]. Improving hardware or data processing algorithms is not what is at the core of our project. However, the Sense-IT is first in providing a new, stable platform that can be considered a *new type of intervention in mental health practice*.

Regarding the OS for which the app was built, the app is currently only available on devices running Android and Wear OS. We did not develop a version for any other OS, such as iOS, because in this study, the hardware was provided to the participants. In addition, Android roughly has had a 75% market share worldwide over the last years in contrast to 22% market share for iOS [51].

More participants could have been included to ensure saturation of feedback and overall group representativeness. To gather a relatively high number of relevant remarks and comments without too many duplicates, we included 3 to 5 users per group per iteration. On the basis of the literature, these numbers seem proportionate, although more could have been better [52,53].

Regarding the selected use case scenario, it could be considered a limitation of this study that it is exclusively focused on the use of the app with patients with BPD *and* low emotional awareness. Of course, use of the app by other patient and nonpatient groups with low emotional awareness seems feasible after context-dependent tailoring [54]. However, if an app works for one of the most challenging groups of users in terms of emotional regulation, it could very well work well for others too [55]. Since the start of the project, researchers from 2 other Dutch health care institutions have joined our group and set up studies with the Sense-IT app within their own settings and their specific patient groups. These studies concern the usefulness, usability, and effect on clinical outcome measures of Sense-IT for patients in forensic psychiatric care with aggression regulation problems, and for adolescents in residential care who have many conflicts because they struggle to detect increasing levels of stress.

Considerable effort was made to ensure *trustworthiness* of this qualitative research project—as, for example, discussed by Shenton [56]. We believe this to be one of the strengths of this study. Shenton mentions 4 criteria for trustworthiness that were originally formulated by Guba for assessing trustworthiness in naturalistic inquiries [57]: credibility (in preference to internal validity), transferability (in preference to external validity/generalizability), dependability (in preference to reliability), and confirmability (in preference to objectivity). To ensure *credibility*, we adopted a *design science* paradigm [58,59] to construct and simultaneously test a scientifically informed approach in designing an mHealth app, using well-recognized research methods. We developed early familiarity with the setting, patients, and therapists. We used different methods and different types of informants. We stimulated honesty in interviewing the participants and used UCD experts to also assess the app. We previously published a detailed report on how this was done and provided a description of the backgrounds, qualifications, and experience of the researchers [28]. To ensure *transferability*, we have provided ample background data to establish the context of study and gave a detailed description of the phenomenon of interest. As mentioned in this section, we started collaborations with other researchers from other settings to study the app in different environments. To ensure *dependability*, we applied several overlapping methods and used an iterative design approach when testing with the patients in which the results of the second cycle served as a *test* of the correct interpretation of the results of the first one. In addition, with the introduction of the EMP framework, we delivered an in-depth methodological description that should allow others to repeat our study. *Confirmability* should be evident from this and previous publication, in which our work was put up for thorough peer review.

What this study added to the literature is an example of how development of an mHealth app within a clinical mental health setting can be challenging, yet feasible, and that it can result in a stable working prototype of an app. It also shows how the use of the multiple user groups is of added value during design and

realization. In this study, the app was not yet tested for clinical effectiveness. Although this is perhaps not so much a limitation of the study, as the usability and stability of the app should be tested before using and testing it as a clinical intervention, it is a question that has to be addressed before use of the app can really be recommended for use as an adjunct in psychotherapies. Such is planned later this year in the same setting as this study took place. Furthermore, a second, graphically more advanced GUI for Sense-IT is currently developed by a dedicated graphic/UX designer to further optimize the UX.

Conclusions

In this study, mHealth development within a clinical mental health setting proved to be challenging, yet feasible and welcomed by targeted users. The Sense-IT app was met with enthusiasm and openness by both patients with BPD and therapists, groups that are both known to be reluctant to embrace technological innovations. The use of the EMP framework and the involvement of multiple user groups proved to be of added value during design and realization, as evidenced by the complementary requirements and perspectives. If the app proves to be effective after further clinical testing, Sense-IT would be one of the first broadly applicable technological interventions in the treatment of BPD—and probably in general mental health care—that is actually *new* as it is not the next form of *talking cure* (ie, psychotherapy), medical treatment, or traditional skills or behavioral training. It would support the treatment of BPD by directly addressing one of the most important factors in BPD, namely limited emotional awareness [6]. In general, it could enable patients to take *therapy* out of the therapist's office into their lives far easier.

Acknowledgments

The authors would like to thank L de Bruin, MSc; M Bout, MSc; and B Loos, BSc, for their assistance in the various cycles of development and their valuable input. They would also like to thank their Sense-IT project partners for their help, funding, and further development and study of the Sense-IT: Prof A Popma, MD, PhD, Institute of Criminal Law and Criminology, Leiden University/De Bascule, Academic Centre for Child and Youth Psychiatry; JF ter Harmsel, MSc, Arkin, Forensic Care; K Nijhof, PhD, Pluryn/Radboud University; and M van Loon—van der Logt, MSc, Pluryn.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of all usability problems and/or needs that were identified.
[PDF File (Adobe PDF File)194 KB - mhealth_v7i10e13479_app1.pdf]

Multimedia Appendix 2
Questionnaire used with the patients after the first iteration.
[PDF File (Adobe PDF File)1467 KB - mhealth_v7i10e13479_app2.pdf]

Multimedia Appendix 3
Questionnaire used with the patients after the second iteration.
[PDF File (Adobe PDF File)594 KB - mhealth_v7i10e13479_app3.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Publishing; 2013.
2. Brickman LJ, Ammerman BA, Look AE, Berman ME, McCloskey MS. The relationship between non-suicidal self-injury and borderline personality disorder symptoms in a college sample. Borderline Personal Disord Emot Dysregul 2014;1:14 [FREE Full text] [doi: 10.1186/2051-6673-1-14] [Medline: 26401298]
3. Blum N, Pfohl B, John DS, Monahan P, Black DW. STEPPS: a cognitive-behavioral systems-based group treatment for outpatients with borderline personality disorder--a preliminary report. Compr Psychiatry 2002;43(4):301-310. [doi: 10.1053/comp.2002.33497] [Medline: 12107867]
4. Linehan MM. Skills Training Manual for Treating Borderline Personality Disorder. New York: Guilford Press; 1993.
5. Lieb K, Zanarini MC, Schmahl C, Linehan MM, Bohus M. Borderline personality disorder. Lancet 2004;364(9432):453-461. [doi: 10.1016/S0140-6736(04)16770-6] [Medline: 15288745]
6. Farrell JM, Shaw IA. Emotional awareness training: a prerequisite to effective cognitive-behavioral treatment of borderline personality disorder. Cogn Behav Pract 1994 Jun;1(1):71-91. [doi: 10.1016/s1077-7229(05)80087-2]
7. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess 2004 Mar;26(1):41-54. [doi: 10.1023/b:joba.0000007455.08539.94]

8. Sloan DM, Kring AM. Measuring changes in emotion during psychotherapy: conceptual and methodological issues. Clin Psychol Sci Pract 2007 Dec;14(4):307-322. [doi: 10.1111/j.1468-2850.2007.00092.x]
9. McMain S, Links PS, Guimond T, Wnuk S, Eynan R, Bergmans Y, et al. An exploratory study of the relationship between changes in emotion and cognitive processes and treatment outcome in borderline personality disorder. Psychother Res 2013;23(6):658-673. [doi: 10.1080/10503307.2013.838653] [Medline: 24156526]
10. Sleuwaegen E, Houben M, Claes L, Berens A, Sabbe B. The relationship between non-suicidal self-injury and alexithymia in borderline personality disorder: 'actions instead of words'. Compr Psychiatry 2017 Aug;77:80-88. [doi: 10.1016/j.comppsych.2017.06.006] [Medline: 28646684]
11. Derks YP, Westerhof GJ, Bohlmeijer ET. A meta-analysis on the association between emotional awareness and borderline personality pathology. J Pers Disord 2017 Jun;31(3):362-384. [doi: 10.1521/pedi_2016_30_257] [Medline: 27387060]
12. Suvak MK, Litz BT, Sloan DM, Zanarini MC, Barrett LF, Hofmann SG. Emotional granularity and borderline personality disorder. J Abnorm Psychol 2011 May;120(2):414-426 [FREE Full text] [doi: 10.1037/a0021808] [Medline: 21171723]
13. Füstös J, Gramann K, Herbert BM, Pollatos O. On the embodiment of emotion regulation: interoceptive awareness facilitates reappraisal. Soc Cogn Affect Neurosci 2013 Dec;8(8):911-917 [FREE Full text] [doi: 10.1093/scan/nss089] [Medline: 22933520]
14. Damasio AR. The Feeling of What Happens: Body and Emotion in the Making of Consciousness. Fort Worth, TX, US: Mariner Books; 2019.
15. Edwards ER, Shivaji S, Wupperman P. The emotion mapping activity: preliminary evaluation of a mindfulness-informed exercise to improve emotion labeling in alexithymic persons. Scand J Psychol 2018 Jun;59(3):319-327. [doi: 10.1111/sjop.12438] [Medline: 29516501]
16. Pollatos O, Werner NS, Duschek S, Schandry R, Matthias E, Traut-Mattausch E, et al. Differential effects of alexithymia subscales on autonomic reactivity and anxiety during social stress. J Psychosom Res 2011 Jun;70(6):525-533. [doi: 10.1016/j.jpsychores.2010.12.003] [Medline: 21624575]
17. Eastabrook JM, Lanteigne DM, Hollenstein T. Decoupling between physiological, self-reported, and expressed emotional responses in alexithymia. Pers Individ Dif 2013 Nov;55(8):978-982. [doi: 10.1016/j.paid.2013.08.001]
18. Stone L, Nielson KA. Intact physiological response to arousal with impaired emotional recognition in alexithymia. Psychother Psychosom 2001;70(2):92-102. [doi: 10.1159/000056232] [Medline: 11244390]
19. Lane RD, Weihs KL, Herring A, Hishaw A, Smith R. Affective agnosia: expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy. Neurosci Biobehav Rev 2015 Aug;55:594-611. [doi: 10.1016/j.neubiorev.2015.06.007] [Medline: 26054794]
20. Preece D, Becerra R, Allan A, Robinson K, Dandy J. Establishing the theoretical components of alexithymia via factor analysis: introduction and validation of the attention-appraisal model of alexithymia. Pers Individ Dif 2017 Dec;119:341-352. [doi: 10.1016/j.paid.2017.08.003]
21. Smith R, Killgore WD, Lane RD. The structure of emotional experience and its relation to trait emotional awareness: a theoretical review. Emotion 2018 Aug;18(5):670-692. [doi: 10.1037/emo0000376] [Medline: 29172623]
22. Taylor GJ, Bagby RM, Parker JD. What's in the name 'alexithymia'? A commentary on 'affective agnosia: expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy'. Neurosci Biobehav Rev 2016 Sep;68:1006-1020. [doi: 10.1016/j.neubiorev.2016.05.025] [Medline: 27235080]
23. Lane RD, Garfield DA. Becoming aware of feelings: integration of cognitive-developmental, neuroscientific, and psychoanalytic perspectives. Neuropsychoanalysis 2005 Jan;7(1):5-30. [doi: 10.1080/15294145.2005.10773468]
24. Cameron K, Ogrodniczuk J, Hadjipavlou G. Changes in alexithymia following psychological intervention: a review. Harv Rev Psychiatry 2014;22(3):162-178. [doi: 10.1097/HRP.0000000000000036] [Medline: 24736520]
25. Samur D, Tops M, Schlinkert C, Quirin M, Cuijpers P, Koole SL. Four decades of research on alexithymia: moving toward clinical applications. Front Psychol 2013;4:861 [FREE Full text] [doi: 10.3389/fpsyg.2013.00861] [Medline: 24312069]
26. Vanheule S, Verhaeghe P, Desmet M. In search of a framework for the treatment of alexithymia. Psychol Psychother 2011 Mar;84(1):84-97; discussion 98. [doi: 10.1348/147608310X520139] [Medline: 22903833]
27. Joyce AS, Fujiwara E, Cristall M, Ruddy C, Ogrodniczuk JS. Clinical correlates of alexithymia among patients with personality disorder. Psychother Res 2013;23(6):690-704. [doi: 10.1080/10503307.2013.803628] [Medline: 23731378]
28. Derks YPMJ, de Visser TD, Bohlmeijer ET, Noordzij ML. mHealth in mental health: how to efficiently and scientifically create an ambulatory biofeedback e-coaching app for patients with borderline personality disorder. Int J Hum Factor Ergon 2017;5(1):61. [doi: 10.1504/ijhfe.2017.10009438]
29. Taelman J, Vandeput S, Spaepen A, van Huffel S. Influence of Mental Stress on Heart Rate and Heart Rate Variability. In: Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering. 2008 Presented at: ECIFMBE'08; November 23-27, 2008; Antwerp, Belgium p. 1366-1369. [doi: 10.1007/978-3-540-89208-3_324]
30. Kreibig SD. Autonomic nervous system activity in emotion: a review. Biol Psychol 2010 Jul;84(3):394-421. [doi: 10.1016/j.biopsycho.2010.03.010] [Medline: 20371374]
31. 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]

32. Garrett JJ. The Elements of User Experience: User-Centered Design for the Web and Beyond - Voices That Matter. New York: New Riders; 2002.
33. Kip H, van Gemert-Pijnen LJ. Holistic development of ehealth technology. In: van Gemert-Pijnen LJ, Sanderman R, Kelders SM, Kip H, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. Abingdon, United Kingdom: Routledge; 2018:151-186.
34. van Gemert-Pijnen LJ, Kip H. Holistic development of ehealth technology. In: van Gemert-Pijnen LJ, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. Abingdon, United Kingdom: Routledge; 2018.
35. Linehan MM, Dexter-Mazza ET. Dialectical behavior therapy for borderline personality disorder. In: Barlow DH, editor. Clinical Handbook of Psychological Disorders: A Step-by-Step Treatment Manual. Fourth Edition. New York, US: Guilford Press; 2008:365-420.
36. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
37. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation In Industry. London: CRC Press; 1996:189-194.
38. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Int 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
39. Brooke J. SUS - a quick and dirty usability scale. In: Jordon PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London: CRC Press; 1996:189-194.
40. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
41. 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]
42. Black DW, Blum N. Systems Training for Emotional Predictability and Problem Solving for Borderline Personality Disorder: Implementing STEPPS Around the Globe. Oxford: Oxford University Press; 2017.
43. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
44. Ogden P. Emotion, mindfulness, and movement: expanding the regulatory boundaries of the window of affect tolerance. In: Fosha D, Siegel DJ, Solomon M, editors. The Healing Power of Emotion: Affective Neuroscience, Development & Clinical Practice. New York: W W Norton & Co; 2009:204-231.
45. Morris ME, Aguilera A. Mobile, social, and wearable computing and the evolution of psychological practice. Prof Psychol Res Pr 2012 Dec;43(6):622-626 [FREE Full text] [doi: 10.1037/a0029041] [Medline: 25587207]
46. Boyer EW, Fletcher R, Fay RJ, Smelson D, Ziedonis D, Picard RW. Preliminary efforts directed toward the detection of craving of illicit substances: the iHeal project. J Med Toxicol 2012 Mar;8(1):5-9 [FREE Full text] [doi: 10.1007/s13181-011-0200-4] [Medline: 22311668]
47. Noordzij ML, Scholten P, Laroy-Noordzij ME. Measuring Electrodermal Activity of Both Individuals With Severe Mental Disabilities and Their Caretakers During Episodes of Challenging Behavior. In: Proceedings of Measuring Behavior 2012, 8th International Conference on Methods and Techniques in Behavioral Research. 2012 Presented at: Measuring Behavior'12; August 28-31, 2012; Utrecht, Netherlands p. 201-205 URL: https://research.utwente.nl/en/publications/measuring-electrodermal-activity-of-both-individuals-with-severe-
48. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018;9:743 [FREE Full text] [doi: 10.3389/fphys.2018.00743] [Medline: 30002629]
49. Karlen W, Kobayashi K, Ansermino JM, Dumont GA. Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation. Physiol Meas 2012 Oct;33(10):1617-1629. [doi: 10.1088/0967-3334/33/10/1617] [Medline: 22986287]
50. van Lier HG, Pieterse ME, Garde A, Postel MG, de Haan HA, Vollenbroek-Hutten MM, et al. A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. Behav Res Methods 2019 Jul 09:- Epub ahead of print. [doi: 10.3758/s13428-019-01263-9] [Medline: 31290128]
51. Statcounter. URL: http://gs.statcounter.com/os-market-share/mobile/worldwide# [accessed 2019-08-28]
52. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Hum-Comput St 1994 Sep;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]
53. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 2009;5(1):34-45 [FREE Full text]
54. Zorzella KP, Muller RT, Cribbie RA, Bambrah V, Classen CC. The role of alexithymia in trauma therapy outcomes: examining improvements in PTSD, dissociation, and interpersonal problems. Psychol Trauma 2019 Jan 28:- (epub ahead of print). [doi: 10.1037/tra0000433] [Medline: 30688506]

55. Stephanidis C, Antona M, Savidis A, Partarakis N, Doulgeraki K, Leonidis A. Design for all: computer-assisted design of user interface adaptation. In: Salvendy G, editor. Handbook of Human Factors and Ergonomics. Hoboken, New Jersey: John Wiley & Sons; 2012:1484-1507.
56. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. Educ Inf 2004 Jul 19;22(2):63-75. [doi: 10.3233/EFI-2004-22201]
57. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ Commun Technol 2012 Dec 20;29(2):75-91 [FREE Full text]
58. March ST, Smith GF. Design and natural science research on information technology. Decis Support Syst 1995 Dec;15(4):251-266. [doi: 10.1016/0167-9236(94)00041-2]
59. van Aken JE. Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological Rules. J Manag Stud 2001 Feb;41(2):- [FREE Full text] [doi: 10.1111/joms.2004.41.issue-2]

Abbreviations

API: application programming interface
BPD: borderline personality disorder
CBT: cognitive behavior therapy
DBT: dialectical behavioral treatment
EDA: electrodermal activity
GUI: graphical user interface
HR: heart rate
HRV: heart rate variability
mHealth: mobile health
OS: operating system
PCEA: physiological correlate of emotional arousal
PPG: photoplethysmogram
SUS: system usability scale
UCD: user-centered design
UX: user experience

Edited by G Eysenbach; submitted 04.02.19; peer-reviewed by E Edwards, S Helweg-Joergensen, A Daros; comments to author 28.02.19; revised version received 13.05.19; accepted 28.07.19; published 15.10.19.

Please cite as:
Derks YPMJ, Klaassen R, Westerhof GJ, Bohlmeijer ET, Noordzij ML
Development of an Ambulatory Biofeedback App to Enhance Emotional Awareness in Patients with Borderline Personality Disorder: Multicycle Usability Testing Study
JMIR Mhealth Uhealth 2019;7(10):e13479
URL: https://mhealth.jmir.org/2019/10/e13479
doi:10.2196/13479
PMID:31617851

Original Paper

A Smart Mobile Health Tool Versus a Paper Action Plan to Support Self-Management of Chronic Obstructive Pulmonary Disease Exacerbations: Randomized Controlled Trial

Lonneke Boer[1], MSc; Erik Bischoff[1], MD, PhD; Maarten van der Heijden[2], PhD; Peter Lucas[2], MD, PhD; Reinier Akkermans[1], IR; Jan Vercoulen[3], PhD; Yvonne Heijdra[4], MD, PhD; Willem Assendelft[1], MD, PhD; Tjard Schermer[1,5], PhD

[1]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[2]Institute for Computing and Information Science, Radboud University, Nijmegen, Netherlands
[3]Department of Medical Psychology, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[4]Department of Pulmonary Diseases, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[5]Netherlands Institute for Health Services Research (NIVEL), Utrecht, Netherlands

Corresponding Author:
Erik Bischoff, MD, PhD
Department of Primary and Community Care
Radboud Institute for Health Sciences
Radboud University Medical Center
PO Box 9101
Nijmegen, 6500 HB
Netherlands
Phone: 31 24 36 65068
Email: erik.bischoff@radboudumc.nl

Abstract

Background: Many patients with chronic obstructive pulmonary disease (COPD) suffer from exacerbations, a worsening of their respiratory symptoms that warrants medical treatment. Exacerbations are often poorly recognized or managed by patients, leading to increased disease burden and health care costs.

Objective: This study aimed to examine the effects of a smart mobile health (mHealth) tool that supports COPD patients in the self-management of exacerbations by providing predictions of early exacerbation onset and timely treatment advice without the interference of health care professionals.

Methods: In a multicenter, 2-arm randomized controlled trial with 12-months follow-up, patients with COPD used the smart mHealth tool (intervention group) or a paper action plan (control group) when they experienced worsening of respiratory symptoms. For our primary outcome exacerbation-free time, expressed as weeks without exacerbation, we used an automated telephone questionnaire system to measure weekly respiratory symptoms and treatment actions. Secondary outcomes were health status, self-efficacy, self-management behavior, health care utilization, and usability. For our analyses, we used negative binomial regression, multilevel logistic regression, and generalized estimating equation regression models.

Results: Of the 87 patients with COPD recruited from primary and secondary care centers, 43 were randomized to the intervention group. We found no statistically significant differences between the intervention group and the control group in exacerbation-free weeks (mean 30.6, SD 13.3 vs mean 28.0, SD 14.8 weeks, respectively; rate ratio 1.21; 95% CI 0.77-1.91) or in health status, self-efficacy, self-management behavior, and health care utilization. Patients using the mHealth tool valued it as a more supportive tool than patients using the paper action plan. Patients considered the usability of the mHealth tool as good.

Conclusions: This study did not show beneficial effects of a smart mHealth tool on exacerbation-free time, health status, self-efficacy, self-management behavior, and health care utilization in patients with COPD compared with the use of a paper action plan. Participants were positive about the supportive function and the usability of the mHealth tool. mHealth may be a valuable alternative for COPD patients who prefer a digital tool instead of a paper action plan.

Trial Registration: ClinicalTrials.gov NCT02553096; https://clinicaltrials.gov/ct2/show/NCT02553096.

(JMIR Mhealth Uhealth 2019;7(10):e14408) doi:10.2196/14408

KEYWORDS

COPD; symptom flare up; mHealth; self-management

Introduction

Exacerbations in chronic obstructive pulmonary disease (COPD) are acute events of transient worsening of the respiratory condition. Exacerbations considerably affect patients' health status [1,2], accelerate the decline in lung function [3], and contribute to COPD-related costs [4]. Despite the substantial impact that exacerbations may have, patients with COPD often have difficulty in recognizing symptom deterioration [5] and do not respond timely or adequately in the course of symptom worsening [6].

Self-management strategies, such as the use of a written exacerbation action plan, have been shown to improve exacerbation outcomes, that is, decrease exacerbation duration [7,8], reduce hospital admissions [9,10], and decrease the impact on health status [10,11]. However, many patients do not adhere to the self-management instructions in their action plans when an exacerbation is imminent [7,12] and thus do not get the benefit of the favorable health effects of timely detection and subsequent intervention.

Telemonitoring, in which patients record and send information on symptoms or physiological measurements to a supervising clinician, may be an alternative approach to self-management strategies to reduce the impact of COPD exacerbations. Beneficial effects have been reported on the number of hospital admissions [13], emergency visits [13], and quality of life in some studies [14,15] but not in all [16]. There is much heterogeneity between telemonitoring interventions regarding devices and clinical content, and the amount of additional support that patients receive. However, in contrast with self-management, telemonitoring strongly depends on the judgement of the clinician and the patient is not expected to interpret his or her own symptoms and signs. Therefore, we have developed [17] and validated [18] an innovative mobile health (mHealth) tool, called the Adaptive Computerized COPD Exacerbation Self-management Support system. This tool aims to tailor self-management support more efficiently and continuously than with a written action plan but without heavily increasing the involvement of health care professionals to monitor input, as is the case with telemonitoring. The mHealth tool integrates information on symptom changes and physiological measurements (ie, pulse oximetry, spirometry, and measurement of body temperature) in an easy-to-use app by means of a mobile phone [17]. On the basis of a decision tree built by a clinical expert panel and a Bayesian prediction model, the tool provides automated, tailored self-management advice to the patient without the involvement of a health care professional [18]. Patients can use the tool at their own initiative to monitor symptom changes at any time of day or night and receive ad hoc, tailored advice.

In this study, we examined the clinical effectiveness of the mHealth tool. We hypothesized that in patients with COPD, the use of the tool would lead to more weeks without exacerbations; improvement in health status, self-efficacy, and self-management behavior; and a reduction in health care utilization compared with the use of a paper exacerbation action plan. We also evaluated how patients valued the tool's supportive function and usability.

Methods

Study Design

This study was a multicenter, parallel, 2-arm, randomized controlled trial with a follow-up of 12 months per patient. After signing informed consent, patients with COPD recruited from general practices and outpatient clinics followed a 20-min self-management educational group session primarily addressing early recognition and prompt treatment of exacerbations. Subsequently, patients were randomized to either (1) usual care according to current COPD guidelines, that is, exacerbation self-management support through the use of a paper exacerbation action plan (control group) or (2) exacerbation self-management support through the use of the mHealth tool (intervention group). Participants in the control group were provided with a written action plan if they did not have one. Participants in the intervention group were instructed not to use their action plan if they had one before participation. This study has been registered at ClinicalTrials.gov (Identifier: NCT02553096) and has been approved by the medical ethics review board, region Arnhem-Nijmegen, the Netherlands (file 2014-1270).

Setting and Participants

Patients were recruited between June 2015 and July 2016 at the pulmonary outpatient clinics of 3 Dutch hospitals and 9 general practices in the city of Nijmegen and the surroundings in the Netherlands. All participating centers delivered care according to the current Dutch COPD guideline [19] and had a pulmonary or practice nurse available for COPD care. Patients were eligible for participation if they (1) were aged at least 40 years, (2) had a spirometry-confirmed diagnosis of COPD (postbronchodilator forced expiratory volume in 1 second (FEV_1)/forced vital capacity<0.7), and (3) had experienced 2 or more symptom-based exacerbations in the previous 12 months, defined as a change for greater than or equal to 2 consecutive days in either greater than or equal to 2 major symptoms (dyspnea, sputum purulence, and sputum amount) or any 1 major symptom plus greater than or equal to 1 minor symptoms (colds, wheeze, sore throat, and cough) [3,20]. Exclusion criteria were (1) severe comorbid conditions that prohibited safe participation, (2) insufficient knowledge of the Dutch language, and (3) persisting difficulties in using the mHealth system after a 2-week practice period and additional assistance.

Randomization Procedure

We used a computer-generated 2-block randomization procedure, stratifying for the health care center. All patients from the participating centers who met the inclusion criteria received a questionnaire from their health care professional with questions related to exacerbations in the previous 12 months. Patients who responded and had experienced 2 or more symptom-based exacerbations in the previous 12 months (see inclusion criteria) were invited by the research team to participate in this study. The allocation order was determined by the order in which eligible patients responded to our invitation to participate (kept by the research assistant). Participants were assigned to one of the groups after signing informed consent during the group meeting by the researcher (LB). Owing to the type of intervention, patients and health care professionals could not be blinded for group assignment. In addition, the research team could not be blinded as it was responsible for the personalization and technical support of the mHealth tool. The study statistician (RA) who was responsible for analyzing the data was blinded for study assignment of the participants until the analyses had been finished.

Intervention and Control Group

Before randomization and after signing the written informed consent, all participants received a 20-min educational session based on the Dutch version of the Living Well with COPD self-management program provided by the nurse in groups of 4 to 10 participants to establish a homogeneous baseline in exacerbation self-management knowledge [12].

Participants in the intervention group were instructed to visit the nurse within 2 weeks after allocation for instructions on the use of the mHealth tool. The tool consisted of a mobile phone (provided by the research team), a pulse oximeter (CMS50D, Contec Medical Systems,), a spirometer (PiKo-1 monitor, nSpire), and a forehead thermometer (FTN, Medisana AG). Patients answered 12 yes-or-no questions concerning changes in symptoms, physical limitations, and emotions using the touch screen on the mobile phone complemented by measurements with the pulse oximeter, spirometer, and forehead thermometer (see Multimedia Appendix 1) [17]. All questions had to be answered to proceed. On the basis of a built-in Bayesian network decision model, the mHealth tool then provided one or more of the following advices: (1) increase your bronchodilator use (including a personalized medication instruction), (2) use your breathing techniques, (3) use your coughing techniques, (4) be thoughtful of how you distribute your energy during the day, (4) contact your health care professional today, (5) measure again tomorrow. Completing the questions and measurements took approximately 5 min. The mHealth tool has been developed in close collaboration with COPD patients and health care professionals [17] and has shown high sensitivity and specificity [18].

Before the trial started, participants in the intervention group were instructed to use the system daily for 2 weeks to get familiarized with the app, mobile phone, spirometer, pulse oximeter, and forehead thermometer. Data were sent to a secured Web-based interface and were monitored by the research team to make sure participants practiced sufficiently. After this 2-week run-in period, the nurses evaluated patients' use of the system, including the physiological measurements. Reference values for each patient's FEV_1 and peripheral oxygen saturation were set. Then, the 12-month follow-up period started. Patients were instructed to use the tool every time they experienced or had any doubts about any change in symptoms or disease burden.

Participants in the control group visited the nurse within 2 weeks of allocation for instructions on the use of a paper exacerbation action plan. When patients did not already possess a written action plan at that moment, the nurses provided the action plan of the Living Well with COPD program [12]. The plan consisted of instructions regarding the self-management of an exacerbation, for example, increase the use of bronchodilators, initiate a standing prescription for a course of prednisolone and/or antibiotic treatment if applicable, or contact the health care professional within 3 days of symptom aggravation.

At the 3-month follow-up, patients in both the intervention and control groups were invited by their nurse to evaluate their self-management of COPD exacerbations. In the intervention group, only the nurses received the patients' entries in the mHealth tool from the research team to enable tailoring of feedback on self-management behavior. In the control group, the nurses evaluated the use of the paper action plan. Patients in both the groups did not receive any feedback on self-management behavior before or after this nurse contact. All patients in both the intervention and control groups continued to have complete access to their health care professionals during the follow-up.

Outcomes and Follow-Up

Our primary outcome was the difference in the number of exacerbation-free weeks between the intervention and control groups. An exacerbation-free week was defined as a week in which there had not been episodes of 2 or more consecutive days with worsening of 2 major symptoms (ie, dyspnea, sputum purulence, and sputum amount) or 1 major and 1 or more minor symptoms (ie, colds, wheeze, sore throat, and cough) [21]. Symptom changes were assessed using the Telephonic Exacerbation Assessment System (TEXAS, Radboudumc), an automated telephone call system that contacted participants weekly on the day and time of their preference [22]. TEXAS consisted of closed questions regarding changes in respiratory symptoms, use of health care resources, and use of respiratory medication in the week before the call, and its validity has been demonstrated previously [22]. Owing to the discontinuation of the contract with the provider of TEXAS, the last 19 participants in the trial received a weekly online questionnaire containing the same questions as TEXAS. These participants used both measuring tools during 2 weeks before stopping with TEXAS, which enabled us to compare data entries from TEXAS with the online survey tool. We found no differences in the data entries.

The secondary outcomes included the following:

- Exacerbation-related outcomes, that is, the number of unscheduled health care contacts, the number of exacerbations treated with antibiotics and/or prednisolone,

and the number of exacerbation-related hospital admissions, all retrieved from patients' medical records, and the number of symptom-based exacerbations as assessed with TEXAS.
- Exacerbation-related self-management behavior, measured with TEXAS or the online questionnaire, and defined as taking 1 or more of the following 3 actions during symptom-based exacerbations: (1) contacting the health care professional, (2) starting a course of prednisolone and/or antibiotics, or (3) maximizing bronchodilator use. We also assessed the time between the date of exacerbation onset and the date of 1 of these 3 actions, defining actions taken within 3 days of exacerbation onset as adherence to the instructions.
- Exacerbation-related self-efficacy, measured with an exacerbation-related self-efficacy scale containing 5 questions. This questionnaire was created for the purpose of this study as, to our knowledge, no questionnaire existed that measured exacerbation-related self-efficacy. Reliability analyses showed a Cronbach alpha of .69 at baseline and .81 at follow-up.
- Health status, measured with (1) the Nijmegen Clinical Screening Instrument (NCSI), which is a battery of instruments measuring 8 subdomains of health status—subjective symptoms, dyspnea emotions, fatigue, behavioral impairment, subjective impairment, general quality of life (QoL), health-related QoL, and satisfaction with relationships [23]; (2) the Clinical Chronic Obstructive Pulmonary Disease Questionnaire (CCQ), which measures 3 subdomains, that is, symptoms, functional status, and mental status, resulting in a total score (24); and (3) the EuroQol-5 dimensions (EQ-5d), [24], which measures health-related quality of life, with a total score based on weighted scores on mobility, self-care, usual activities, pain/discomfort, and anxiety/depression as well as a vertical Visual Analogue Scale varying between 0 and 100.

At the start and at 12 months, data were gathered on exacerbation history, self-efficacy, and health status. CCQ and EQ-5d were also completed at 3, 6, and 9 months of follow-up.

At 12 months, information on health care utilization, lung function, respiratory medication use, and comorbid conditions was extracted from the participants' medical records. In addition, all participants were asked to evaluate the supportive function of either the mHealth tool or the paper action plan by using a paper survey including closed-ended questions regarding the use, difficulty in use, and intended future use of the mHealth tool or the paper action plan. Besides, 3 questions were asked related to clarity, suitability, and follow-up of the advice given by the mHealth tool or the paper action plan. All questions included answers on a 7-point rating scale from strongly disagree (score 1) to strongly agree (score 7). The survey also included 1 question about frequency of usage at times of symptom worsening, with answers on a 7-point rating scale varying from 1=never to 7=always. In addition, participants of the intervention group were asked to complete the System Usability Scale (SUS) [25]. The SUS contains 10 questions on system usability, which are calculated into 1 total score between 0 and 100. SUS scores less than 68 are considered as low, greater than or equal to 68 and less than or equal to 80.3 as good, and greater than 80.3 as excellent.

Sample Size Calculation and Statistical Analyses

Sample size calculation using analysis of variance showed that we needed 43 participants in each group for 80% power (alpha=.05, 2 sided) to detect an increase of 6 exacerbation-free weeks per year and anticipating a dropout rate of 20% (9/43). The calculation was based on a previous dataset [12] in which we found a mean of 44 exacerbation-free weeks, with SD 4.5 weeks. We used all available data from all participants based on the intention-to-treat principle. Missing data were not imputed.

We used the data recorded by the secured Web-based interface to assess the actual usage of the mHealth tool. We analyzed the answers to the paper evaluation survey to assess participants' self-reported use of the mHealth tool and the paper action plan. Negative binomial regression analyses, controlling for follow-up time per participant, age and gender were used to analyze our primary outcome, that is, the number of exacerbation-free weeks, as well as the number of unscheduled health care contacts, self-reported exacerbations, exacerbations treated with antibiotics and/or prednisolone, and exacerbation-related hospital admissions.

To test the effect of the mHealth tool on the rate of symptom-based exacerbations and self-management behavior, we extracted exacerbation episodes from the TEXAS database. Each new episode was preceded by at least 2 exacerbation-free weeks or 2 weeks with missing data [22]. To assess patient delay in taking action when an exacerbation was imminent, the numbers of days were calculated between the date of exacerbation onset and the following actions: (1) the first date of contact with a health care professional, (2) starting date of the course of prednisolone and/or antibiotics, or (3) date of increase of bronchodilators. We categorized these variables into 2 groups: less than 3 days (according to instructions), and greater than or equal to 3 days. Separate multilevel logistic regression analyses were performed taking into account the clustering effect of exacerbations within patients and controlling for age and gender to examine (1) whether the mHealth tool led to higher percentages of self-management actions in case of an exacerbation compared with the paper action plan and (2) whether these actions were more often taken timely by patients in the intervention group compared with the control group.

We used generalized estimating equation regression analyses to estimate the effect of the mHealth tool on changes between baseline and follow-up scores of the self-efficacy scale, NCSI, CCQ, and EQ-5d compared with the paper action plan. We analyzed the CCQ and EQ-5d with all 5 measurement time points. We used 2-tailed *t* tests for independent samples and chi-square test to analyze the differences in the participants' preferences between the mHealth tool and the paper action plan.

Statistical significance was assumed at $P<.05$ based on 2-sided tests. We used IBM SPSS Statistics 25 for the analyses.

Results

Patient Characteristics

Of the 87 patients included in the study, 43 were randomized to the intervention group. In addition, 45 patients were recruited from the hospitals and 42 from the general practices. Among them, 13% (11/87) dropped out of the study, 16% (7/43) in the intervention, and 9% (4/44) in the control group. A flowchart of the participants in the study is shown in Figure 1. Demographic and baseline characteristics are shown in Table 1.

Mean duration of follow-up was 48.1 (SD 11.7) weeks, and 11 COPD-related hospital admissions (6 in the intervention group and 5 in the control group) were reported as serious adverse events to the medical ethics review board.

Usage of Mobile Health Tool and Paper Action Plan

From the Web-based interface, it appeared that 38 of the 43 patients (88%) in the mHealth group used the app 727 times in total during follow-up. No data on usage was available for 5 patients. The range in frequency of usage was 1 to 250 times with a median of 7 (25%-75% interquartile range was 3-14). Results of the evaluation questionnaire showed that more patients reported to have used their mHealth tool often (scores 6 and 7 on the 7-point rating scale) compared with patients in the control group who reported to have used their paper action plan (44.4% vs 17.2%, respectively).

Figure 1. Flow diagram of the participants through the study. mHealth: mobile health.

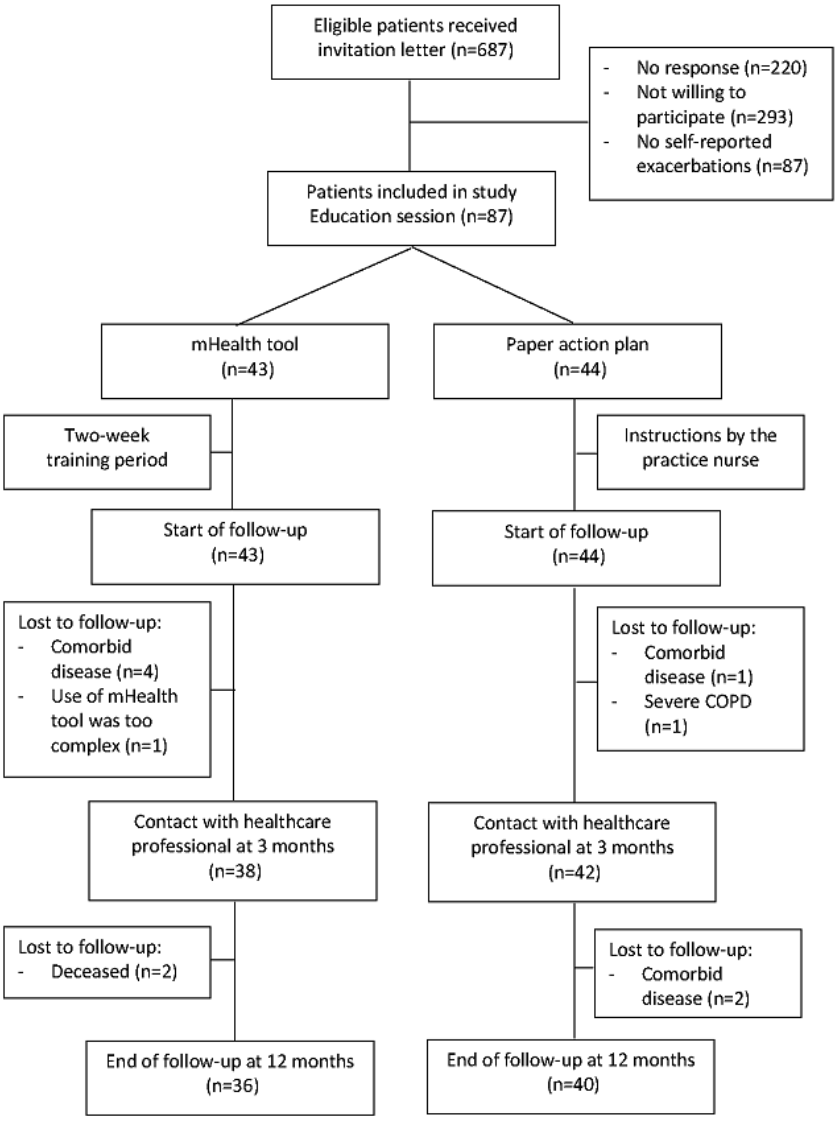


Table 1. Baseline and demographic characteristics of the study population (N=87) per treatment arm.

Characteristic	Mobile health tool intervention group (n=43)	Paper action plan control group (n=44)
Recruited in hospital, n (%)	21 (49)	24 (55)
Follow-up in weeks, mean (SD)	48.3 (12.6)	49.8 (10.9)
Age (years), mean (SD)	69.3 (8.8)	65.9 (8.9)
Male sex, n (%)	25 (58)	29 (66)
Postbronchodilator FEV_1[a] (% predicted), mean (SD)	53.0 (21.5)	52.1 (19.8)
Medical Research Council dyspnea score, mean (SD)	2.5 (1.2)	2.6 (1.3)
Currently smoking, n (%)	13 (30)	11 (25)
Use of paper action plan prior to follow-up, n (%)	11 (26)	17 (39)
Respiratory drug treatment, n (%)		
Long-acting bronchodilators	27 (63)	26 (59)
Short-acting bronchodilators	30 (70)	31 (71)
Inhaled corticosteroids (ICS)	7 (16)	13 (30)
Long-acting bronchodilators + ICS	22 (51)	24 (55)
Low education level, n (%)	19 (44)	17 (39)
Diagnosis of COPD >5 years, n (%)	29 (68)	28 (64)
Currently working, n (%)	6 (14)	10 (23)
Relevant comorbidities, n (%)		
Joint disorders	13 (30)	13 (30)
Cardiac disorders	12 (28)	12 (27)
Back pain	8 (19)	14 (32)
Diabetes	3 (7)	3 (7)
Depression and/or anxiety	3 (7)	2 (5)

[a]FEV_1: forced expiratory volume in 1 second.

Table 2. Comparison of exacerbation-related outcomes between intervention group and control group (N=85).

Outcome	Mobile health tool intervention group (n=41), mean (SD)	Paper action plan control group (n=44), mean (SD)	Rate ratio (95% CI)[a]	*P* value
Exacerbation-free weeks[b]	30.6 (13.3)	28.0 (14.8)	1.21 (0.77-1.90)	.40
Unscheduled health care consultations because of respiratory complaints[c]	1.6 (1.7)	1.6 (2.0)	0.89 (0.50-1.60)	.70
Symptom-based exacerbations[b]	4.5 (2.3)	4.3 (2.1)	1.07 (0.65-1.75)	.80
Exacerbations treated with antibiotics and/or prednisolone[c]	1.1 (1.5)	1.0 (1.3)	1.01 (0.53-1.93)	.97
Exacerbation-related hospital admissions[c]	0.15 (0.43)	0.14 (0.41)	1.25 (0.35-4.44)	.74

[a]Calculated using negative binomial regression analyses, relative to participants' follow-up time, controlling for age and gender.

[b]Data retrieved from weekly patient reports.

[c]Data retrieved from patient medical files.

Exacerbation-Free Time and Other Exacerbation-Related Outcomes

Patients in the intervention group did not differ statistically significantly from patients in the control group in the number of weeks without exacerbations (mean 30.6 weeks, SD 13.3 weeks vs mean 28.0 weeks, SD 14.8 weeks, respectively; rate ratio [RR] 1.21; 95% CI 0.77-1.91; see Table 2). In addition, no statistically significant differences were found in the number of symptom-based exacerbations (mean 4.5, SD 2.3 vs mean 4.3, SD 2.1, respectively; RR 1.07, 95% CI 0.65-1.75), and the number of exacerbations treated with antibiotics and/or prednisolone (mean 1.0, SD 1.5 vs mean 1.0, SD 1.3, respectively; RR 1.01, 95% CI 0.53-1.93). Furthermore, no

statistically significant between-group differences were found in the number of unscheduled health care contacts (mean 1.6, SD 1.7 vs mean 1.6, SD 2.0, respectively; RR 0.89, 95% CI 0.50-1.60) and exacerbation-related hospital admissions (mean 0.15, SD 0.43 vs mean 0.14, SD 0.41, respectively; RR 1.25, 95% CI 0.35-4.44).

Self-Management Behavior

A total of 377 symptom-based exacerbation episodes were identified. Table 3 shows that there were no statistically significant differences between the intervention and control groups in the frequency of contacting a health care professional (odds ratio [OR] 0.94, 95% CI 0.51-1.73), initiating a course of antibiotics and/or prednisolone (OR 1.16, 95% CI 0.55-2.44), and increasing bronchodilator use (OR 1.08, 95% CI 0.56-2.06). We found no differences between the intervention and control groups in timely action, that is, contacting a health care professional (OR 2.21, 95% CI 0.78-6.23), starting a course of antibiotics and/or prednisolone (OR 1.46, 95% CI 0.48-4.42), or increasing bronchodilator use (OR 1.15, 95% CI 0.61-2.16) within 3 days of exacerbation onset (Table 4).

Self-Efficacy and Health Status

We found no statistically significant difference in exacerbation-related self-efficacy between the intervention and control groups when comparing baseline scores with 12-month follow-up scores. In addition, there were no differences between the groups in changes between baseline and follow-up scores of the subscales of the NCSI, CCQ and EQ-5d (Table 5).

Participant Evaluation of the Self-Management Support Tools

A total of 58 (67%) participants returned an evaluation form, of which 28 were in the intervention group. The mHealth tool was rated as a more useful support tool than the paper action plan (*P*=.02). No differences were found between the mHealth tool and the action plan in the self-reported frequency of use; in difficulty and future use of the tool; or in clarity, suitability, and follow-up of the advice. Overall, 26 participants of the intervention group completed all 10 questions of the SUS. The mean score was 78.5 (SD 14.4).

Table 3. Self-reported self-management behavior during exacerbation onset (N=377 exacerbations).

Self-management action	Mobile health tool intervention group (n=187), n (%)	Paper action plan control group (n=190), n (%)	Odds ratio (95% CI)[a]	*P* value
Contact health care professional	61 (32.6)	68 (35.8)	0.94 (0.51-1.73)	.83
Start prednisolone and/or antibiotics	64 (34.2)	62 (32.6)	1.16 (0.55-2.44)	.69
Increase bronchodilator use	135 (72.2)	135 (71.1)	1.08 (0.56-2.06)	.82

[a]Calculated using multilevel logistic regression analyses, including participant as cluster variable, controlling for age and gender.

Table 4. Self-reported self-management behavior within 3 days of exacerbation onset.

Self-management action	Mobile health tool intervention group		Paper action plan control group		Odds ratio (95% CI)[a]	*P* value
	n	<3 days, n (%)	n	<3 days, n (%)		
Contact health care professional	55	20 (36.4)	63	17 (27.0)	2.21 (0.78-6.23)	.13
Start prednisolone and/or antibiotics	58	23 (39.7)	57	23 (40.4)	1.46 (0.48-4.42)	.50
Increase bronchodilator use	122	87 (71.3)	117	80 (68.4)	1.15 (0.61-2.16)	.67

[a]Calculated using multilevel logistic regression analyses, including participant as cluster variable, controlling for age and gender.

Table 5. Baseline and follow-up scores of exacerbation-related self-efficacy and measures of health status.

Outcome	Mobile health tool intervention group		Paper action plan control group			*P* value
	Baseline (n=43), mean (SD)	12-month (n=35), mean (SD)	Baseline (n=44), mean (SD)	12-month (n=41), mean (SD)	Beta (95% CI)[a]	
Exacerbation-related self-efficacy[b]	2.91 (0.43)	2.98 (0.41)	2.84 (0.41)	2.87 (0.52)	0.03 (−0.17 to 0.22)	.91
Health status measurements[c]						
NCSI[d] QOL[e]	12.90 (8.32)	13.01 (8.27)	19.08 (11.93)	17.11 (12.14)	2.53 (−1.28 to −6.33)	.19
NCSI HRQOL[f]	4.19 (1.74)	4.00 (1.61)	4.68 (1.78)	4.76 (1.93)	−0.16 (−0.89 to −0.57)	.66
NCSI relationship	2.42 (0.93)	2.58 (0.87)	3.39 (1.74)	3.24 (1.59)	0.30 (−0.26 to −0.85)	.29
NCSI subjective impairment	11.84 (5.80)	11.20 (4.12)	14.41 (6.57)	13.22 (6.57)	0.74 (−1.38 to −2.85)	.5
NCSI behavioral impairment	22.11 (17.89)	20.50 (15.58)	19.11 (17.28)	20.43 (21.76)	−1.77 (−7.20 to −3.66)	.52
NCSI subjective symptoms	9.86 (4.88)	9.23 (4.39)	11.91 (4.86)	11.05 (4.79)	0.35 (−1.40 to −2.11)	.69
NCSI dyspnea emotions	8.77 (2.42)	8.71 (2.86)	11.59 (3.92)	10.73 (4.32)	0.85 (−0.58 to −2.27)	.24
NCSI fatigue	35.93 (10.96)	35.23 (9.45)	37.32 (10.17)	37.73 (10.20)	−1.80 (−5.43 to −1.84)	.33
CCQ[g] total	2.06 (1.02)	1.84 (0.77)	2.31 (1.09)	2.16 (1.05)	−0.06 (−0.38 to −0.26)	.7
CCQ symptoms	2.41 (1.12)	2.16 (0.80)	2.60 (1.27)	2.49 (1.24)	−0.22 (−0.67 to −0.23)	.34
CCQ functional status	2.22 (1.38)	2.03 (1.21)	2.53 (1.36)	2.41 (1.32)	0.05 (−0.30 to −0.40)	.76
CCQ mental status	1.03 (1.01)	0.81 (0.79)	1.30 (0.99)	1.00 (1.05)	0.09 (−0.34 to −0.53)	.68
EQ-5d[h]	0.81 (0.15)	0.79 (0.16)	0.74 (0.20)	0.77 (0.21)	−0.05 (−0.13 to −0.03)	.22
EQ VAS[i]	65.53 (17.37)	70.94 (12.92)	64.20 (15.35)	62.63 (19.14)	6.28 (−0.56 to −13.11)	.07

[a]Beta (B) indicates the difference between the intervention and control groups on differences between baseline and at 12 months.

[b]Higher score is positive.

[c]Higher score is negative.

[d]NCSI: Nijmegen Clinical Screening Instrument.

[e]QOL: quality of life.

[f]HRQOL: health-related quality of life.

[g]CCQ: Clinical Chronic Obstructive Pulmonary Disease Questionnaire.

[h]EQ-5d: EuroQol-5 dimensions.

[i]EQ VAS: EuroQol Visual Analogue Scale.

Discussion

Principal Findings

In this study, we examined the clinical effects of a smart mHealth tool to support COPD patients in the detection and treatment of exacerbations without the interference of a health care professional. Our primary hypothesis that the use of mHealth would lead to more weeks without exacerbations than care as usual, that is, the use of a paper action plan, was not confirmed. In addition, we did not find differences in exacerbation frequency, health care utilization, or self-management behavior between patients who used the mHealth tool and patients who used the paper action plan. Furthermore, patients using the tool did not report higher exacerbation-related self-efficacy or better health status than patients using a paper action plan. Patients evaluated the usability of the mHealth tool as good and considered it as more supportive than the action plan.

Comparison With Previous Work

So far, studies on the effects of electronic health (eHealth) in the management of COPD have focused on the use of telemonitoring apps [26,27]. However, the mHealth tool in our study critically differs from telemonitoring tools as it enables the patient to be solely responsible for initiating treatment, without the interference of a health care professional. This impedes comparing our results with those found in telemonitoring trials. A recent Cochrane review showed that, until now, there is only limited evidence that interventions aimed at facilitating, supporting, and sustaining self - management in patients with COPD and delivered via smart technology may improve outcomes such as health status and physical activity [28]. We found no effects of the mHealth tool on health status, measured with generic questionnaires, such as the NCSI and the EQ-5D, or with a disease-specific questionnaire, such as the CCQ.

In this study, we could not demonstrate positive effects on exacerbation-free weeks, exacerbation frequency, and health

care utilization. Our primary outcome, *exacerbation-free weeks*, is directly related to exacerbation recovery time and may better reflect the burden of exacerbations in patients with COPD than exacerbation frequency does [21]. The mean number of exacerbation-free weeks we found during the 1-year follow-up period was 30.6 weeks in the intervention group and 28.0 weeks in the control group. These mean values are in line with the 33.4 weeks that we found in a previous study in which we examined the relationship between exacerbation frequency and exacerbation-free time in a cohort of 166 COPD patients [21]. These values suggest that there was enough room for improvement as the participants seem to suffer from symptom worsening in approximately 20 weeks in the 1-year follow-up period. Although patients rated the mHealth tool as a more useful support tool than the paper action plan, we found no differences between the mHealth tool and the paper action plan on self-management behavior and exacerbation-related self-efficacy. Previous studies showed that patients with COPD were able to use mHealth apps, including reporting daily symptoms and measuring physiological variables [29], and were able to interpret clinical data and use these within their self-management approach regardless of previous knowledge [30]. These findings are in line with our finding that participants in the intervention group rated the usability of the mHealth tool as good on average.

Limitations

The major strength of this study is that we used a well-designed and validated mHealth tool to support self-management behavior [18]. In a qualitative study, Korpershoek et al demonstrated that to optimize engagement, mHealth interventions should be attractive, rewarding, safe, and tailored to the patient needs [31]. Our mHealth tool has been developed using feedback from patients with COPD, and its treatment advice can be tailored to the individual patient. To further optimize the use of the tool, patients were familiarized with the technique during a 2-week run-in period.

However, this study also has limitations. Two important limitations may have led to the statistically nonsignificant results of our primary outcome exacerbation-free weeks. First, we were surprised that the mean exacerbation-free time and its standard deviation at the 12-month follow-up differed substantially from the study data on which our power calculation was based [12]. We cannot explain this. As a result, the sample size in this study may have been too small to actually detect any statistically significant differences in our primary outcome. Conversely, when we performed our sample size calculation, there were no other studies available to provide data on exacerbation-free time. Second, offering both the intervention and control groups a short education session on the recognition and treatment of exacerbations before the start of the study and providing a paper action plan to the control group may have reduced the room for improvement and may have diluted potential differences between the 2 groups on our primary outcome exacerbation-free weeks. The purpose of the education session was to equalize the level of self-management knowledge among all participants and between both groups before the start of the trial. Our choice may have upgraded self-management knowledge and skills of all participants, although we did not measure what the participants had actually learned from the session to verify this assumption. We chose to provide our control group with a paper action plan according to the recommendations in the current national COPD guideline [19]; thereby supporting the self-management knowledge and skills of the control group participants. Many general practitioners and chest physicians in the Netherlands have not (yet) integrated the use of paper action plans in their daily practice. This was also found in our study, in which only 25.6% participants of the intervention group and 38.6% participants of the control group used a paper action plan before the study. Conversely, previous use of a paper action plan could have affected our primary outcome exacerbation-free weeks if patients in the intervention group continued to use the paper action plan instead of using the mHealth tool. Although we cannot refute this assumption with our data, we think that this effect (if any) has been small as we found that all intervention group participants used the mHealth tool and that more patients in the intervention group used the mHealth tool than patients in the control group used their paper plans.

There are also other limitations. Of the 467 patients that responded to the study invitation and met the inclusion criteria, 283 were not willing to participate. This may have led to selection bias. We believe that the risk of contamination because of the individual randomization procedure is negligible, as both the mHealth tool and the paper action plan were used at home, outside the reach of the health care professional. Although in the last 4 months of the trial we had to switch to collecting exacerbation-related outcomes through a digital survey tool instead of the automated TEXAS telephone calls [22], this did not impact the rate of data entry, and the participants experienced no difficulties. So, although it was an undesirable deviation from the protocol, we have no concerns on the reliability of the primary outcome data. Finally, 12 months of follow-up may have been too short for the mHealth tool to reach its maximum effect, as it may take more time to sustain engagement with the technology and change self-management behavior [28].

Future Research

Although we were not able to demonstrate positive effects of our mHealth tool, we still believe that the use of eHealth tools, including machine learning techniques, better suits the goals of patient-centered care and self-management support than telemonitoring tools where a health care professional monitors the patient from a distance. Besides, in this study, the patients using the mHealth tool evaluated it as usable and more supportive than patients using a conventional supportive tool, that is, the written paper action plan. Future research should focus more on patients who are specifically interested in using digital tools in their daily life, as these patients may have greater benefit from them [28]. More information on patient perceptions of the use of the mHealth tool is also needed and could be collected by patient interviews or open-ended surveys. In addition, more research is needed on the clinical effects of the mHealth tool when used appropriately and the factors that are associated with appropriate use. However, for these purposes, there is a need for studies with larger patient populations than in this study.

Conclusions

In this study, we examined the effectiveness of an mHealth tool designed to support COPD patients in their self-management of symptom worsening to reduce the impact of exacerbations. The app was not designed to replace the health care professional but to reduce patient delay. Patients evaluated the app's usability as good and as more supportive than the paper action plan. Although this study did not show beneficial effects of the mHealth app compared with the use of a paper action plan, based on patient's preference, it may be a valuable alternative to a paper action plan in the management of COPD.

Acknowledgments

The authors would like to thank all patients who participated in this study for their commitment to the study as well as all team members, especially the nurses and physicians of the participating hospitals and general practices.

Authors' Contributions

All authors contributed to the design of the study. LB, EB, and JV supervised the collection of the data. LB and RA performed the statistical analysis. LB led the writing of the paper, which was coled by EB, TS, and WA and assisted by all others. All authors had full access to all of the study, assisted in the interpretation of the data, and have seen and approved the final version of the paper.

Conflicts of Interest

None of the authors received any support from any company for the submitted work. LB, MvdH, RA, YH, JV, and WA have nothing to disclose. PL has a patent 20140206949 issued to Petrus Lucas, EB received personal fees from Boehringer Ingelheim and GlaxoSmithKline outside the submitted work, and TS received a grant from GlaxoSmithKline and personal fees from Boehringer Ingelheim outside the submitted work.

Multimedia Appendix 1
Contents of the mobile health tool.
[PDF File (Adobe PDF File)42 KB - mhealth_v7i10e14408_app1.pdf]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File)1709 KB - mhealth_v7i10e14408_app2.pdf]

References

1. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1998 May;157(5 Pt 1):1418-1422. [doi: 10.1164/ajrccm.157.5.9709032] [Medline: 9603117]
2. Langsetmo L, Platt RW, Ernst P, Bourbeau J. Underreporting exacerbation of chronic obstructive pulmonary disease in a longitudinal cohort. Am J Respir Crit Care Med 2008 Mar 15;177(4):396-401. [doi: 10.1164/rccm.200708-1290OC] [Medline: 18048806]
3. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002 Oct;57(10):847-852 [FREE Full text] [doi: 10.1136/thorax.57.10.847] [Medline: 12324669]
4. Wedzicha JA, Seemungal TA. COPD exacerbations: defining their cause and prevention. Lancet 2007 Sep 1;370(9589):786-796. [doi: 10.1016/S0140-6736(07)61382-8] [Medline: 17765528]
5. Kessler R, Ståhl E, Vogelmeier C, Haughney J, Trudeau E, Löfdahl CG, et al. Patient understanding, detection, and experience of COPD exacerbations: an observational, interview-based study. Chest 2006 Jul;130(1):133-142. [doi: 10.1378/chest.130.1.133] [Medline: 16840393]
6. Wilkinson TM, Donaldson GC, Hurst JR, Seemungal TA, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004 Jun 15;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
7. Bischoff EW, Hamd DH, Sedeno M, Benedetti A, Schermer TR, Bernard S, et al. Effects of written action plan adherence on COPD exacerbation recovery. Thorax 2011 Jan;66(1):26-31. [doi: 10.1136/thx.2009.127621] [Medline: 21037270]
8. Zwerink M, Kerstjens HA, van der Palen J, van der Valk P, Brusse-Keizer M, Zielhuis G, et al. (Cost-)effectiveness of self-treatment of exacerbations in patients with COPD: 2 years follow-up of a RCT. Respirology 2016 Apr;21(3):497-503 [FREE Full text] [doi: 10.1111/resp.12697] [Medline: 26647766]
9. Sedeno MF, Nault D, Hamd DH, Bourbeau J. A self-management education program including an action plan for acute COPD exacerbations. COPD 2009 Oct;6(5):352-358. [doi: 10.1080/15412550903150252] [Medline: 19863364]

10. Lenferink A, Brusse-Keizer M, van der Valk PD, Frith PA, Zwerink M, Monninkhof EM, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Aug 4;8:CD011682 [FREE Full text] [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]
11. Trappenburg JC, Monninkhof EM, Bourbeau J, Troosters T, Schrijvers AJ, Verheij TJ, et al. Effect of an action plan with ongoing support by a case manager on exacerbation-related outcome in patients with COPD: a multicentre randomised controlled trial. Thorax 2011 Nov;66(11):977-984. [doi: 10.1136/thoraxjnl-2011-200071] [Medline: 21785156]
12. Bischoff EW, Akkermans R, Bourbeau J, van Weel C, Vercoulen JH, Schermer TR. Comprehensive self management and routine monitoring in chronic obstructive pulmonary disease patients in general practice: randomised controlled trial. Br Med J 2012 Nov 28;345:e7642 [FREE Full text] [doi: 10.1136/bmj.e7642] [Medline: 23190905]
13. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: an overview of reviews. Chron Respir Dis 2017 Aug;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]
14. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease: Cochrane review and meta-analysis. Br J Gen Pract 2012 Nov;62(604):e739-e749 [FREE Full text] [doi: 10.3399/bjgp12X658269] [Medline: 23211177]
15. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-127. [doi: 10.1258/jtt.2009.090812] [Medline: 20197355]
16. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. Br Med J 2013 Oct 17;347:f6070 [FREE Full text] [doi: 10.1136/bmj.f6070] [Medline: 24136634]
17. van der Heijden M, Lucas PJ, Lijnse B, Heijdra YF, Schermer TR. An autonomous mobile system for the management of COPD. J Biomed Inform 2013 Jun;46(3):458-469 [FREE Full text] [doi: 10.1016/j.jbi.2013.03.003] [Medline: 23500485]
18. Boer LM, van der Heijden M, van Kuijk NM, Lucas PJ, Vercoulen JH, Assendelft WJ, et al. Validation of ACCESS: an automated tool to support self-management of COPD exacerbations. Int J Chron Obstruct Pulmon Dis 2018;13:3255-3267 [FREE Full text] [doi: 10.2147/COPD.S167272] [Medline: 30349231]
19. Long Alliantie Nederland. 2016. Zorgstandaard COPD URL: http://www.longalliantie.nl/zorgstandaard-copd/ [accessed 2019-08-29]
20. Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2000 May;161(5):1608-1613. [doi: 10.1164/ajrccm.161.5.9908022] [Medline: 10806163]
21. Boer LM, Bischoff EW, Borgijink X, Vercoulen JH, Akkermans RP, Kerstjens HA, et al. 'Exacerbation-free time' to assess the impact of exacerbations in patients with chronic obstructive pulmonary disease (COPD): a prospective observational study. NPJ Prim Care Respir Med 2018 Apr 3;28(1):12 [FREE Full text] [doi: 10.1038/s41533-018-0079-5] [Medline: 29615628]
22. Bischoff EW, Boer LM, Molema J, Akkermans R, van Weel C, Vercoulen JH, et al. Validity of an automated telephonic system to assess COPD exacerbation rates. Eur Respir J 2012 May;39(5):1090-1096 [FREE Full text] [doi: 10.1183/09031936.00057811] [Medline: 21920893]
23. Peters JB, Daudey L, Heijdra YF, Molema J, Dekhuijzen PN, Vercoulen JH. Development of a battery of instruments for detailed measurement of health status in patients with COPD in routine care: the Nijmegen clinical screening instrument. Qual Life Res 2009 Sep;18(7):901-912 [FREE Full text] [doi: 10.1007/s11136-009-9502-2] [Medline: 19543807]
24. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
25. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London: Taylor & Francis; 1996:189-194.
26. Segrelles Calvo G, Gómez-Suárez C, Soriano JB, Zamora E, Gónzalez-Gamarra A, González-Béjar M, et al. A home telehealth program for patients with severe COPD: the PROMETE study. Respir Med 2014 Mar;108(3):453-462 [FREE Full text] [doi: 10.1016/j.rmed.2013.12.003] [Medline: 24433744]
27. Farmer A, Williams V, Velardo C, Shah SA, Yu L, Rutter H, et al. Self-management support using a digital health system compared with usual care for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2017 May 3;19(5):e144 [FREE Full text] [doi: 10.2196/jmir.7116] [Medline: 28468749]
28. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 May 23;5:CD011425 [FREE Full text] [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
29. Hardinge M, Rutter H, Velardo C, Shah SA, Williams V, Tarassenko L, et al. Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study. BMC Med Inform Decis Mak 2015 Jun 18;15:46 [FREE Full text] [doi: 10.1186/s12911-015-0171-5] [Medline: 26084626]

30. Williams V, Price J, Hardinge M, Tarassenko L, Farmer A. Using a mobile health application to support self-management in COPD: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e392-e400 [FREE Full text] [doi: 10.3399/bjgp14X680473] [Medline: 24982491]
31. Korpershoek Y, Bos-Touwen ID, de Man-van Ginkel JM, Lammers J, Schuurmans MJ, Trappenburg J. Determinants of activation for self-management in patients with COPD. Int J Chron Obstruct Pulmon Dis 2016;11:1757-1766 [FREE Full text] [doi: 10.2147/COPD.S109016] [Medline: 27536087]

Abbreviations

CCQ: Clinical Chronic Obstructive Pulmonary Disease Questionnaire
COPD: chronic obstructive pulmonary disease
eHealth: electronic health
EQ-5D: EuroQol-5 dimensions
FEV1: forced expiratory volume in 1 second
mHealth: mobile health
NCSI: Nijmegen Clinical Screening Instrument
OR: odds ratio
QoL: quality of life
RR: rate ratio
SUS: System Usability Scale
TEXAS: Telephonic Exacerbation Assessment System

Edited by G Eysenbach; submitted 17.04.19; peer-reviewed by A Nguyen, P Ware; comments to author 06.06.19; revised version received 25.07.19; accepted 18.08.19; published 09.10.19.

Please cite as:
Boer L, Bischoff E, van der Heijden M, Lucas P, Akkermans R, Vercoulen J, Heijdra Y, Assendelft W, Schermer T
A Smart Mobile Health Tool Versus a Paper Action Plan to Support Self-Management of Chronic Obstructive Pulmonary Disease Exacerbations: Randomized Controlled Trial
JMIR Mhealth Uhealth 2019;7(10):e14408
URL: https://mhealth.jmir.org/2019/10/e14408
doi:10.2196/14408
PMID:31599729

Original Paper

Evaluation of Mobile Apps Targeted at Patients With Spondyloarthritis for Disease Monitoring: Systematic App Search

Yu Heng Kwan[1*], PhD; Wei Jie Ong[2,3*], BBus, BAcc; Mengfei Xiong[4], MBBS; Ying Ying Leung[2,3,5], MBChB, MD; Jie Kie Phang[2], BSc; Charmaine Tze May Wang[2], MB BCh BAO; Warren Fong[2,3,5], MBBS, MRCP

[1]Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

[2]Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore

[3]Duke-NUS Medical School, Singapore, Singapore

[4]Department of Family Medicine, SingHealth, Singapore, Singapore

[5]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Yu Heng Kwan, PhD
Program in Health Services and Systems Research
Duke-NUS Medical School
8 College Road
Singapore, 169857
Singapore
Phone: 65 90231226
Email: yuheng@u.duke.nus.edu

Abstract

Background: There are many apps developed for patients with spondyloarthritis in the market, but their purpose and quality are not objectively evaluated.

Objective: The objective of this study was to identify and evaluate existing publicly available, high-quality apps that use validated measurement instruments for monitoring spondyloarthritis disease activity.

Methods: We conducted a review of apps available on the Apple App Store and the Google Play Store based on a combination of keywords and inclusion and exclusion criteria. Validated disease activity measurement instruments were identified. Data regarding app characteristics, including the presence of validated disease activity measurement, were extracted. The Mobile App Rating Scale (MARS) was used to review the apps for user experience.

Results: A total of 1253 apps were identified in the app stores, and 5 apps met the criteria and were further analyzed. Moreover, 2 apps (MySpA and Group for Research and Assessment of Psoriasis and Psoriatic Arthritis App) contained some of the validated disease activity monitoring instruments for specific spondyloarthritis subtypes. These 2 apps were also rated good on the MARS (with total mean scores ≥4 out of 5), whereas the other apps scored poorly in comparison.

Conclusions: There are 2 high-quality spondyloarthritis disease activity monitoring apps publicly available, but they only target 2 spondyloarthritis subtypes—ankylosing spondylitis and psoriatic arthritis. There is a lack of high-quality apps that can measure disease activity for other spondyloarthritis subtypes, and no app that consolidates all validated disease activity instruments across subtypes was available.

(JMIR Mhealth Uhealth 2019;7(10):e14753) doi:10.2196/14753

KEYWORDS

spondylarthropathies; mobile apps; ankylosing spondylitis; psoriatic arthritis

Introduction

Background

Spondyloarthritis (SpA) is a heterogeneous group of chronic inflammatory diseases with interrelated clinical features and genetic linkages and includes ankylosing spondylitis (AS), psoriatic arthritis (PsA), inflammatory bowel disease–associated arthritis, reactive arthritis, and undifferentiated SpA [1]. These phenotypically diverse and unpredictable diseases are associated with decreased performance of activities of daily living, quality of life, and work productivity [2]. One of the goals for SpA management is to monitor disease activity, which can be used to modify both pharmacological and nonpharmacological treatments [3].

With increased smartphone ownership, mobile health (mHealth) becomes a relevant and fast-growing field of health care delivery, whereby the apps developed for smartphone users could be potential useful tools for both patient self-management and enhanced communications between patients and physicians [4,5]. Apps that targeted the management of chronic diseases had shown benefits for patients in research studies, including conditions such as obesity and diabetes, which led to improved clinical outcomes and maintenance of high-quality medical care [6]. In the field of rheumatology, it had been shown that it was beneficial for patients with rheumatoid arthritis (RA), where management was enhanced with electronic tools such as digitally recorded disease activity (joint counts) and electronic patient-reported outcome measures (ePROMs). These patients with RA who regularly completed ePROMs and digitally tracked their disease had a better compliance to their medications, less activity handicap in daily living, and more positive outlook for their future [7].

There is a need for potential users to be able to objectively evaluate the quality of health care apps as it is unclear for the average user as to which apps offer evidence-based tools and education [6]. It had been shown that many health apps neither adhere to evidence-based guidelines nor involve medical professionals during development [8]. Without a proper evaluation system, users could be trusting apps that are not based on best medical practices or evidence, which could lead to harm. As the availability of such apps increases, it is important that users make educated decisions about the apps they use. For a disease like SpA that has many different subtypes, it has been shown that there are benefits to having a single app that allows patients to monitor their disease activity across all subtypes [2,9].

Objectives

The objective of this study was to identify existing publicly available, high-quality apps that use validated measurement instruments for monitoring SpA disease activity. The specific aims of this review were to evaluate and determine the features and quality of apps designed to monitor SpA disease activity by (1) identifying and summarizing the available apps and the key disease activity monitoring features, (2) comparing the app features with validated instruments for monitoring of SpA disease activity, and (3) rating app quality according to the Mobile App Rating Scale (MARS) [10]. This will enable users to make informed decisions about specific apps and may identify deficiencies in the mHealth apps for SpA disease activity monitoring currently available for future development purposes.

Methods

This review comprised 4 stages. Stage 1 involved a systematic search of the apps on both the Apple App Store and Google Play Store based on a combination of keywords (stage 1: app identification). This included a screening process in which apps were screened for inclusion into the next stage. Stage 2 involved identifying the validated disease monitoring instruments for various SpA subtypes. Stage 3 involved the extraction and recording of relevant data regarding each app. Finally, the apps were evaluated (stage 4) using the MARS checklist.

App Identification

A systematic search of the Apple App Store and Google Play Store was conducted from July 15, 2018, to July 21, 2018, to identify all potentially relevant apps. The review was conducted following the Preferred Reporting Items for Systematic reviews Meta-Analyses guidelines [11]. Search terms included “spondyloarthritis” OR “ankylosing” OR “ankylosing spondylitis” OR “psoriatic” OR “psoriatic arthritis” OR “reactive arthritis” OR “inflammatory bowel disease related arthritis” OR “arthritis.” The app store description of each identified app was read and compared with the inclusion and exclusion criteria. Apps were included if they were (1) a smartphone-based app, (2) capable of running on Android or iOS operation systems, (3) in English language, (4) useful for people with SpA or to assist people with SpA for their clinical care, and (5) available for download in the app store (Apple App Store or Google Play Store). Apps were excluded if (1) a condition other than SpA was targeted; (2) the app included only treatment algorithms; (3) it was explicitly only for clinician use; or (4) app content was for information, education, or reference only (ie, no data entry). Apps not updated before 2017 were also excluded because of potential incompatibility with newer operating systems that could underrepresent the functionality of the app. When an app was found in both the Google Play Store and Apple App Store, both versions were included so any differences between operating systems could be identified, and the reviewers interacted with both apps.

To ensure the capture of all relevant apps, the search on the Apple App Store (using an iPhone on a Singapore internet protocol [IP] address) was compared with a website [12] (Singapore Apple App Store) using the search term “arthritis.” “Arthritis” was used as the search term as it returned the most apps during the main search.

Validated Spondyloarthritis Disease Activity Monitoring Instruments

As part of our objective was to assess apps that measured disease activity in patients with SpA, these instruments were identified (Table 1) using guidelines from the Assessment of SpondyloArthritis international Society [13] and European League Against Rheumatism [3,14], along with the subtype of SpA they relate to. Apps were then evaluated as to whether they possess the functionality to calculate any of these instruments.

Table 1. Validated disease activity monitoring instruments for spondyloarthritis disease activity monitoring.

Instrument	Spondyloarthritis subtype	Instrument components
Bath Ankylosing Spondylitis Functional Index [13]	AS[a]	PGA[b]
Bath Ankylosing Spondylitis Disease Activity Index [13]	AS	PGA
Ankylosing spondylitis disease activity score [3]	AS	PGA, erythrocyte sedimentation rate, and CRP[c]
Minimal disease activity [3]	PsA[d]	28 tender joint count, 28 swollen joint count, VAS[e], Psoriasis Area Severity Index, body surface area, PGA, Health Assessment Questionnaire, and tender entheseal points
Disease activity index for psoriatic arthritis [3]	PsA	68 tender joint count, 66 swollen joint count, CRP, VAS, and PGA
Psoriatic Arthritis Impact of Disease [14]	PsA	PGA

[a]AS: ankylosing spondylitis.

[b]PGA: patient global assessment (of disease activity).

[c]CRP: C-reactive protein.

[d]PsA: psoriatic arthritis.

[e]VAS: (patient pain) visual analog score.

Data Extraction

The following data about all apps were recorded: app name, platform (Android or iOS), developer, current version, size, cost, number of installs, and user star ratings. General functions of the apps as well as any validated SpA disease activity monitoring instruments were recorded descriptively.

App Rating Using the Mobile App Rating Scale

The MARS was developed as a means to determine the quality and classification of mHealth apps [10]. The MARS identified 23 items that were rated on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent), with a description provided for each anchor rating. The MARS comprises 5 different domains and each contributes to the overall evaluation of the app quality:

1. Engagement (5 items)—the extent to which the app engages the target users;
2. Functionality (4 items)—how easy the app is to learn and navigate and overall app performance;
3. Aesthetics (3 items)—the graphics, visual appeal, and style of the app;
4. Information quality (7 items)—the accuracy of the app description, the quality and quantity of information in the app, and whether the information is verified by scientific evidence;
5. Subjective quality (4 items).

Scoring for the MARS was done by calculating a mean for each category and an overall mean score, which has been proven to have good internal consistency and interrater reliability, so it is a reliable method to rate and compare mobile apps [10,15,16]. Apps that score ≥4 out of 5 on the overall MARS rating are considered *good* [17].

All the apps were rated by 2 independent reviewers (MFX and WJO) using the MARS [10]. Before embarking on the review, the 2 reviewers discussed the use of the MARS in the context of apps for people with SpA. The target group was determined to be "all people with SpA aged 18 years or older with some experience using smartphone apps." The reviewers also considered all items of the MARS and confirmed that all were applicable to SpA and that no additional app-specific items were required [10], as recommended by the developers of the MARS.

Before reviewing all the apps identified in the search, both reviewers assessed and discussed an excluded app to ensure alignment of understanding of the MARS rating criteria. The reviewers then independently rated all apps using the MARS. Each app was used for at least 10 min to gain an adequate understanding of the app functionality. Apps were tested on July 24, 26, and 27, 2018, using an iPhone 8 Plus running on iOS 11.4.1 and an MI NOTE LTE equipped with Android version 6.0. 1 MMB29M, using the app version downloaded on July 24, 2018. Any matters or doubts about specific apps were debated between the 2 reviewers and a third reviewer (YHK) who has significant experience in SpA, and consensus was reached.

Scores were calculated for each MARS item, along with a total mean score. The mean score from 2 reviewers was calculated. No app had been tested in clinical studies. Therefore, item 19 of the MARS, *evidence base*, was excluded from calculations. Interrater reliability of the MARS subscales and total quality score were calculated using the intraclass correlation coefficient (ICC) in STATA SE 14.0 by StataCorp LLC.

Results

Overview

The app selection process detailed in the Methods section is summarized in Figure 1. The characteristics and functions of 5 apps were recorded in Tables 2 and 3. The use of validated SpA disease monitoring activity instruments was also recorded, as shown in Table 4. Their MARS scores ranged from 2.93 to 4.04 (possible range 0.00-5.00), as shown in Table 5.

Figure 1. Flow diagram of systematic app search and selection from the Google Play Store and Apple App Store.

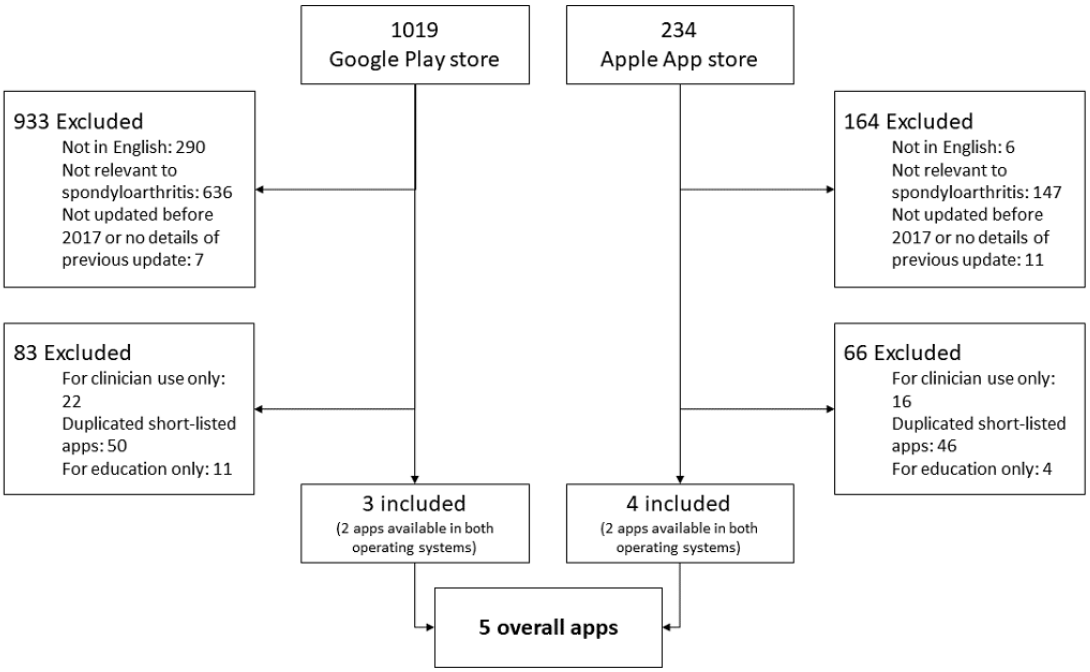


Table 2. Operating system, developer, version, and size of included apps.

App	Operating system	Developer	iOS version	iOS size (MB)	Android version	Android size (MB)
AS Health Storylines	iOS and Android	Self Care Catalysts Inc	7.24	18	5.4.4_10MAY2017_1	12
Psoriasis Manager	iOS	At Point of Care LLC	10.0.0	20.7	—[a]	—
GRAPPA[b] App	iOS and Android	GRAPPA	1.1	30.3	1.1	26
ArthritisPower	iOS	Global Healthy Living Foundation Inc	2.0.6	14.9	—	—
MySpA	Android	Barts Health National Health Service Trust	—	—	1.1.6	80

[a]Not available.

[b]GRAPPA: Group for Research and Assessment of Psoriasis and Psoriatic Arthritis.

Table 3. App description and target users.

App	App description	Target subtype (as per app description)
AS Health Storylines	Input data to monitor disease	Ankylosing spondylitis
Psoriasis Manager	Input data to monitor disease	PsA[a]
Group for Research and Assessment of Psoriasis and Psoriatic Arthritis App	Input data to monitor disease	PsA
ArthritisPower	Input data to monitor disease	PsA
MySpA	General information; exercise plans; input data to monitor disease	Axial spondyloarthritis and PsA

[a]PsA: psoriatic arthritis.

Table 4. App inclusion of the various spondyloarthritis activity measure and component measurement instruments and other functionality of included apps.

App	Data entered	Instruments or laboratory measures	Composite disease activity measure	Allows users to record and retrieve disease activity data on multiple occasions	General functions
AS Health Storylines	Medication and symptoms	PGA[a]	—[b]	History (journal style) and graph (trend of scores)	Symptom and medication tracker
Psoriasis Manager	Medication and symptoms	PGA	—	History (journal style)	Treatment tracker and patient journal
Group for Research and Assessment of Psoriasis and Psoriatic Arthritis App	Symptoms	PGA	Psoriatic Arthritis Impact of Disease and minimal disease activity	—	Disease monitoring information about PsA[c]
ArthritisPower	Medication, symptoms, and laboratory results	PGA	—	History (journal style) and graph (trend of scores)	Symptom and medication tracker
MySpA	Medication, symptoms, and laboratory results	PGA, 66/68 joint count, C-reactive protein, and erythrocyte sedimentation rate	Bath Ankylosing Spondylitis Functional Index and Bath Ankylosing Spondylitis Disease Activity Index	History (journal style)	Information about axial spondyloarthritis and PsA disease monitoring

[a]PGA: patient global assessment (of disease activity).

[b]Not available.

[c]PsA: psoriatic arthritis.

Table 5. Mean Mobile App Rating Scale ratings of included apps.

App name	MARS[a] categories											
	Engagement mean scores (5 items)		Functionality mean scores (4 items)		Aesthetics mean scores (3 items)		Information mean scores (7 items)		Subjective mean scores (4 items)		Overall MARS mean scores	
	Android	iOS	Android	iOS	Android	iOS	Android	iOS	Android	iOS	Android	iOS
AS Health Storylines	2.00	2.00	3.75	3.75	3.67	3.67	3.00	3.00	2.25	2.25	2.93	2.93
Psoriasis Manager	—[b]	2.80	—	4.00	—	3.33	—	3.00	—	2.25	—	3.08
Group for Research and Assessment of Psoriasis and Psoriatic Arthritis App	3.40	3.40	4.50	4.50	4.33	4.33	4.20	4.20	3.75	3.75	4.04	4.04
ArthritisPower	—	2.40	—	4.00	—	3.33	—	3.50	—	2.00	—	3.05
MySpA	3.40	—	4.50	—	4.33	—	4.20	—	3.75	—	4.04	—

[a]MARS: Mobile App Rating Scale.

[b]Not available.

The search retrieved 1019 Android apps from the Google Play Store. Of these, 1016 were excluded, leaving 3 apps for analysis (Figure 1). A total of 234 iOS apps were retrieved from the Apple App Store. After exclusion of 230 apps, 4 apps remained for analysis. A total of 91 apps were found in the concurrent fnd.io search of the Singapore Apple App Store, out of which no relevant further apps were found. As 2 apps were available in both operating systems, a total of 5 different apps were included; all were free apps.

Characteristics and Functions of Included Apps

The information on app platform, developer, version, and size are shown in Table 2. As no app had different function between operating systems, the apps are presented only once in Tables 2-5. The app description and target user (as derived from the app store description) are shown in Table 3. None of the iOS apps included had any star rating. Table 4 shows data entry and main functionality in the apps. All the apps allowed users to enter data, such as symptoms and medication. One app (MySpA) included the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and Bath Ankylosing Spondylitis Functional Index (BASFI) measures, whereas another app (Group for Research and Assessment of Psoriasis and Psoriatic Arthritis [GRAPPA] App) included the Psoriatic Arthritis Impact Of Disease (PsAID) measures and the minimal disease activity (MDA) calculator with the Health Assessment Questionnaire (HAQ) and Psoriasis Area and Severity Index embedded within.

App inclusion of component measurement instruments, composite disease activity measures calculated, and app functionality to record and retrieve data over time are shown in Table 4. A total of 2 apps (GRAPPA App and MySpA) included at least one validated composite measure of SpA disease activity, out of which none provided the formulae for calculation of the composite disease activity measure. Moreover, 4 apps (AS Health Storylines, Psoriasis Manager, ArthritisPower, and MySpA) included a function allowing data (such as patients' medication and symptoms) to be recorded and retrieved. In addition, 1 app (MySpA) included both composite disease activity measure and allowed data recording and retrieval.

Rating of Apps on Mobile App Rating Scale

The MARS ratings for included apps are shown in Table 5. The ICC for the MARS ratings were greater than or equal to 0.91 for all the MARS sections. For overall MARS ratings, the ICC was 0.99 (95% CI 0.98-0.99), confirming good interrater reliability. Of the 2 apps that scored ≥4 out of 5 on the overall MARS rating (considered good [17]), both apps (MySpA and GRAPPA App) included composite disease activity measures that were validated (BASFI and BASDAI for MySpA; MDA and PsAID for GRAPPA App), but only MySpA had a data tracking function. The overall MARS scores for the apps ranged from 2.93 to 4.04, indicating significant variation in the quality of apps. Subjective quality (2.00-3.75) and engagement (2.00-3.40) showed greatest variability.

Discussion

Principal Findings

The aim of this study was to identify existing high-quality apps that used validated measurement instruments for monitoring SpA disease activity. This review of apps showed that a significant proportion of the publicly available apps specifically designed for SpA were for general education purposes only (50% [4/8] for iOS apps and 78% [11/14] for Android apps), whereas there was a smaller category of apps for tracking SpA symptoms and calculation of validated disease activity measures (50% [4/8] for iOS apps and 22% [3/14] for Android apps). Of the symptom-tracking apps, 60% (3/5) did not use validated instruments in this study. We would recommend patients to use apps with validated disease activity tracking instruments and app developers to develop future apps using such instruments as well.

Only 1 app (scoring ≥4 out of 5 on the overall MARS), MySpA, included both a symptom-tracking function and calculation of validated composite measures of AS disease activity (BASFI and BASDAI). Hence, patients with AS wishing to track their symptoms are encouraged to use MySpA. Another app, GRAPPA App (also scoring ≥4 out of 5 on the overall MARS), included the calculation of validated composite measures of PsA disease activity (MDA and PsAID) but lacked a tracking function. People with PsA can opt to use the app for calculating their disease activity scores but may want to record it elsewhere for tracking purposes. Although the app AS Health Storylines does not have the function to calculate a validated disease activity monitoring score, it has the function to track and remind patients to take medications, which can potentially be synchronized with the overall health care ecosystem. App developers can consider adding both the calculation and recording functions to future developments.

Although some disease activity instruments can be calculated easily by patients (via apps) using ePROMs, such as the BASDAI, BASFI, and PsAID, there are others that require either clinician inputs (tender and swollen joint counts and tender entheseal points for MDA) or laboratory test results (C-reactive protein and erythrocyte sedimentation rate for the Ankylosing Spondylitis Disease Activity Score). There are studies showing that self-reported joint counts have shown to reasonably predict clinician-performed joint counts for RA patients [18,19], though this assumption that self-reported joint counts will be sufficient for measurement of SpA disease activity needs to be validated further. GRAPPA App, which has a calculator for MDA, includes an HAQ questionnaire that the patient completes him or herself. It is also worth noting that no app included tracking of disease activity for more than 1 SpA subtype and that apps for tracking disease activity of SpA subtypes other than AS and PsA were not found. There is potential for future app developers to consider filling in these gaps or developing a single app with the capability of disease monitoring across all the different subtypes of SpA. We would also like to highlight that although these apps are useful in helping the patients track their disease activity, appropriate treatment and therapy should still be done in collaboration with their health care teams.

Apps rated have a wide range of the MARS scores (2.93 to 4.04), indicating highly inconsistent quality of apps in terms of user experience. App developers wishing to optimize user experience can consider using the MARS criteria as a checklist, along with collaborating with the key stakeholders such as people with SpA and medical professionals. Item 19 of the MARS, *evidence base*, was excluded from all calculations because no app had been studied in clinical trials [10]. Therefore, clinical trials should be conducted for any future apps developed for SpA disease activity monitoring to determine the clinical impact on outcomes for people with SpA, as well as to ensure cost-effectiveness and to undergo external quality review [20].

This review has limitations. Only apps available in app stores accessed from a Singapore IP address and in English language were included. However, a preliminary search of the iTunes store of the United States with the term "arthritis" suggested that the main search of the app stores had captured all relevant apps in English language. Patients were not included in the rating of the app in this study. Future studies can be performed to address this gap.

App quality was assessed using the MARS. Although the MARS was recently developed and had not been extensively validated, it had now been used in several other app evaluations [15,16,21], and had consistently proven good interrater reliability between reviewers. The ICC was 0.99 for the overall MARS score in this study, confirming good interrater reliability. Caregivers can consider using the MARS criteria for rating apps that they wish to recommend to their patients.

Assessment of data privacy and security was not included in the MARS but was a commonly considered criterion of health

software quality not included in this study [22]. Data privacy and security considerations are of utmost importance and will need to be evaluated against the specific regulatory requirements of the country in which the app is being used and should be included in future studies.

Conclusions

In conclusion, to our knowledge, this is the first review of high-quality apps for monitoring SpA disease activity that use validated measurement instruments. This review indicated that the apps MySpA and GRAPPA App were high-quality apps that used validated disease activity measures, which could assist in the management of SpA subtypes AS and PsA, respectively. However, there is a lack of apps that consolidate the measurement and tracking of disease activity of different SpA subtypes into a single app. Future app development can consider developing apps that bridge these gaps.

Acknowledgments

This study is partially supported by Singapore's Ministry of Health Traditional Chinese Medicine Research Grant (TCMRG/SGH310301).

Conflicts of Interest

None declared.

References

1. Raychaudhuri SP, Deodhar A. The classification and diagnostic criteria of ankylosing spondylitis. J Autoimmun 2014;48-49:128-133. [doi: 10.1016/j.jaut.2014.01.015] [Medline: 24534717]
2. Png K, Kwan YH, Leung YY, Phang JK, Lau JQ, Lim KK, et al. Measurement properties of patient reported outcome measures for spondyloarthritis: a systematic review. Semin Arthritis Rheum 2018 Oct;48(2):274-282. [doi: 10.1016/j.semarthrit.2018.02.016] [Medline: 29573848]
3. Smolen JS, Schöls M, Braun J, Dougados M, FitzGerald O, Gladman DD, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann Rheum Dis 2018 Jan;77(1):3-17 [FREE Full text] [doi: 10.1136/annrheumdis-2017-211734] [Medline: 28684559]
4. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies - Second Global Survey on Ehealth. Geneva, Switzerland: World Health Organization; 2011.
5. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht UV. 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]
6. Luo D, Wang P, Lu F, Elias J, Sparks JA, Lee YC. Mobile apps for individuals with rheumatoid arthritis: a systematic review. J Clin Rheumatol 2019 Apr;25(3):133-141. [doi: 10.1097/RHU.0000000000000800] [Medline: 29933327]
7. El Miedany Y, El Gaafary M, Youssef S, Bahlas S, Almedany S, Ahmed I, et al. Toward electronic health recording: evaluation of electronic patient-reported outcome measures system for remote monitoring of early rheumatoid arthritis. J Rheumatol 2016 Dec;43(12):2106-2112. [doi: 10.3899/jrheum.151421] [Medline: 27633823]
8. Subhi Y, Bube SH, Bojsen SR, Thomsen AS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
9. Kwan YH, Fong W, Tan VI, Lui NL, Malhotra R, Østbye T, et al. A systematic review of quality-of-life domains and items relevant to patients with spondyloarthritis. Semin Arthritis Rheum 2017 Oct;47(2):175-182. [doi: 10.1016/j.semarthrit.2017.04.002] [Medline: 28487138]
10. 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]
11. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
12. Fnd: Experience the App Store and iTunes Anywhere. URL: https://fnd.io/ [accessed 2019-07-18] [WebCite Cache ID 6lo8YnNxq]
13. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The assessment of spondyloarthritis international society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis 2009 Jun;68(Suppl 2):ii1-i44. [doi: 10.1136/ard.2008.104018] [Medline: 19433414]
14. Gossec L, de Wit M, Kiltz U, Braun J, Kalyoncu U, Scrivo R, EULAR PsAID Taskforce. A patient-derived and patient-reported outcome measure for assessing psoriatic arthritis: elaboration and preliminary validation of the psoriatic

arthritis impact of disease (PsAID) questionnaire, a 13-country EULAR initiative. Ann Rheum Dis 2014 Jun;73(6):1012-1019. [doi: 10.1136/annrheumdis-2014-205207] [Medline: 24790067]
15. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for people with rheumatoid arthritis to monitor their disease activity: a review of apps for best practice and quality. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]
16. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
17. Reyes A, Qin P, Brown CA. A standardized review of smartphone applications to promote balance for older adults. Disabil Rehabil 2018 Mar;40(6):690-696. [doi: 10.1080/09638288.2016.1250124] [Medline: 27868438]
18. Pincus T, Yazici Y, Bergman M, Maclean R, Harrington T. A proposed continuous quality improvement approach to assessment and management of patients with rheumatoid arthritis without formal joint counts, based on quantitative routine assessment of patient index data (RAPID) scores on a multidimensional health assessment questionnaire (MDHAQ). Best Pract Res Clin Rheumatol 2007 Aug;21(4):789-804. [doi: 10.1016/j.berh.2007.02.009] [Medline: 17678835]
19. Cheung PP, Gossec L, Mak A, March L. Reliability of joint count assessment in rheumatoid arthritis: a systematic literature review. Semin Arthritis Rheum 2014 Jun;43(6):721-729. [doi: 10.1016/j.semarthrit.2013.11.003] [Medline: 24332117]
20. Buijink AW, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
21. Patel R, Sulzberger L, Li G, Mair J, Morley H, Shing MN, et al. Smartphone apps for weight loss and smoking cessation: quality ranking of 120 apps. N Z Med J 2015 Sep 4;128(1421):73-76. [Medline: 26370762]
22. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]

Abbreviations

AS: ankylosing spondylitis
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Functional Index
CRP: C-reactive protein
ePROM: electronic patient-reported outcome measure
GRAPPA: Group for Research and Assessment of Psoriasis and Psoriatic Arthritis
HAQ: Health Assessment Questionnaire
ICC: intraclass correlation coefficient
IP: internet protocol
MARS: Mobile App Rating Scale
MDA: minimal disease activity
mHealth: mobile health
PGA: patient global assessment
PsA: psoriatic arthritis
PsAID: Psoriatic Arthritis Impact of Disease
RA: rheumatoid arthritis
SpA: spondyloarthritis
VAS: visual analog score

Edited by G Eysenbach; submitted 17.05.19; peer-reviewed by K Munoz Esquivel, P Gill; comments to author 19.08.19; revised version received 26.08.19; accepted 30.08.19; published 28.10.19.

Please cite as:
Kwan YH, Ong WJ, Xiong M, Leung YY, Phang JK, Wang CTM, Fong W
Evaluation of Mobile Apps Targeted at Patients With Spondyloarthritis for Disease Monitoring: Systematic App Search
JMIR Mhealth Uhealth 2019;7(10):e14753
URL: http://mhealth.jmir.org/2019/10/e14753/
doi:10.2196/14753
PMID:31661080

Original Paper

An Ambulatory Blood Pressure Monitor Mobile Health System for Early Warning for Stroke Risk: Longitudinal Observational Study

Guangyu Wang[1], PhD; Silu Zhou[2], MSc; Shahbaz Rezaei[3]; Xin Liu[3], PhD; Anpeng Huang[2], PhD

[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China

[2]National Institute of Health Data Science, Peking University, Beijing, China

[3]Computer Science Department, University of California, Davis, CA, United States

Corresponding Author:
Anpeng Huang, PhD
National Institute of Health Data Science
Peking University
38 Xueyuan Road
Haidian District
Beijing, 100191
China
Phone: 86 1062755832
Email: hapku@pku.edu.cn

Abstract

Background: Stroke, as a leading cause of death around the globe, has become a heavy burden on our society. Studies show that stroke can be predicted and prevented if a person's blood pressure (BP) status is appropriately monitored via an ambulatory blood pressure monitor (ABPM) system. However, currently there exists no efficient and user-friendly ABPM system to provide early warning for stroke risk in real-time. Moreover, most existing ABPM devices measure BP during the deflation of the cuff, which fails to reflect blood pressure accurately.

Objective: In this study, we sought to develop a new ABPM mobile health (mHealth) system that was capable of monitoring blood pressure during inflation and could detect early stroke-risk signals in real-time.

Methods: We designed an ABPM mHealth system that is based on mobile network infrastructure and mobile apps. The proposed system contains two major parts: a new ABPM device in which an inflation-type BP measurement algorithm is embedded, and an abnormal blood pressure data analysis algorithm for stroke-risk prediction services at our health data service center. For evaluation, the ABPM device was first tested using simulated signals and compared with the gold standard of a mercury sphygmomanometer. Then, the performance of our proposed mHealth system was evaluated in an observational study.

Results: The results are presented in two main parts: the device test and the longitudinal observational studies of the presented system. The average measurement error of the new ABPM device with the inflation-type algorithm was less than 0.55 mmHg compared to a reference device using simulated signals. Moreover, the results of correlation coefficients and agreement analyses show that there is a strong linear correlation between our device and the standard mercury sphygmomanometer. In the case of the system observational study, we collected a data set with 88 features, including real-time data, user information, and user records. Our abnormal blood pressure data analysis algorithm achieved the best performance, with an area under the curve of 0.904 for the low risk level, 0.756 for the caution risk level, and 0.912 for the high-risk level. Our system enables a patient to be aware of their risk in real-time, which improves medication adherence with risk self-management.

Conclusions: To our knowledge, this device is the first ABPM device that measures blood pressure during the inflation process and has obtained a government medical license. Device tests and longitudinal observational studies were conducted in Peking University hospitals, and they showed the device's high accuracy for BP measurements, its efficiency in detecting early signs of stroke, and its efficiency at providing an early warning for stroke risk.

(JMIR Mhealth Uhealth 2019;7(10):e14926) doi:10.2196/14926

KEYWORDS

ambulatory blood pressure monitor; mHealth; stroke-risk early warning; abnormal blood pressure data analyzing; longitudinal observational study

Introduction

As a common cerebrovascular disease caused by a hemorrhage or ischemia, stroke has become the primary cause of adult disability, and its incidence is increasing steadily as society ages [1,2]. According to the World Health Organization (WHO), worldwide 1/6 people has a stroke during their lifetime, among which more than one-third of those affected die from their stroke [3]. Worldwide, mortality from stroke and related cardiovascular diseases constitutes around 20% of all deaths, and around 33% of all stroke survivors are affected by depression after [4,5]. It has been estimated that stroke cost in the first half of the 21st century is around USD $1.52 trillion, which is three times as high as the US annual gross tax revenue [6].

To address these challenges, a lot of recent research has focused on stroke prevention. Many risk factors have been shown to be crucial for stroke, including age, gender, hypertension, diabetes, arrhythmia, cholesterol, smoking, and drinking, among others, with hypertension being the most significant controllable factor [7]. In addition, a growing number of studies have demonstrated the predictive value of ambulatory blood pressure monitoring (ABPM) in assessing the risk of first-ever and recurrent stroke [8-10]. According to the American Stroke Association, 77% of patients have a blood pressure (BP) ≥140/90 mmHg when they have a stroke for the first time [11]. Studies have shown that there is a high correlation between Morning Blood Pressure Surge and stroke. As reported by Argentino et al [12], the average probability of stroke incidence from 6 AM to 12 AM is 3.8 times greater than the other times of the day [13]. Furthermore, with each 10-mmHg increase in Morning Blood Pressure Surge, stroke probabilities increase by 44% (P=.004) [14]. As a result, a real-time Morning Blood Pressure Surge monitoring system for stroke risk warnings can dramatically reduce stroke probability.

With the rapid development of wireless Internet technology and the increasing number of mobile phone users, mobile health (mHealth) technology has emerged as a potential solution to health care delivery for people with chronic diseases [15-17]. Altintas et al proposed a wearable, 24-hour, low-stress blood pressure monitor system to conduct less painful and less stressful, 24-hour blood pressure monitoring [18]. Yang et al designed a set of General Packet Radio Service ambulatory blood pressure monitoring systems, which consist of a data acquisition module, main control module, communication module, cloud platform, and intelligent terminal [19]. Tsoi et al constructed an integrative electronic health platform with a BP telemonitoring cloud system in elderly community centers to control hypertension [20]. In a recent study, Wu et al introduced the design and development of an mHealth system for strengthening secondary prevention of stroke in rural China and further evaluated its feasibility through a survey of the dominant user group [21]. Moreover, other BP monitoring systems [22-25] use a mixture of solutions, including a homemade blood pressure measurement device, a smartphone- or PC-based management unit, or Bluetooth or ZigBee for data transmission, which help users to measure and manage their daily blood pressure.

However, these studies utilized BP devices based on deflation-type instead of inflation-type BP measurement algorithms, which do not accurately reflect the real state of blood pressure. In clinical diagnosis, most pathological changes inside blood vessels may not be captured by deflation-type BP measurement [26,27] because the mechanical behavior of the brachial artery cannot be restored to a normal state within a sharp transition from inflation to deflation. In addition, the abnormal signals generated from Morning Blood Pressure Surge monitoring of the existing approaches do not provide early-risk alert of a stroke to patients or relatives in real-time. Driven by this motivation, we designed a new ABPM mHealth system which can prevent a stroke in real-time. This system consists of two major parts: a new ABPM device for BP monitoring and data collection, and an abnormal BP data analysis algorithm at a data center for real-time early risk warning services. Significantly, an inflation-type blood pressure measurement algorithm has been proposed and embedded into the new device, which helps to avoid the elastic deformation of blood vessels and any interference signals during measurement. At the data center, the abnormal BP data analysis algorithm runs to classify stroke-risk levels based on real-time BP data, medical records, and health state information. To evaluate the ABPM mHealth system, a longitudinal observational study was conducted, which demonstrated the accuracy and usefulness of the early stroke-risk warning function.

Methods

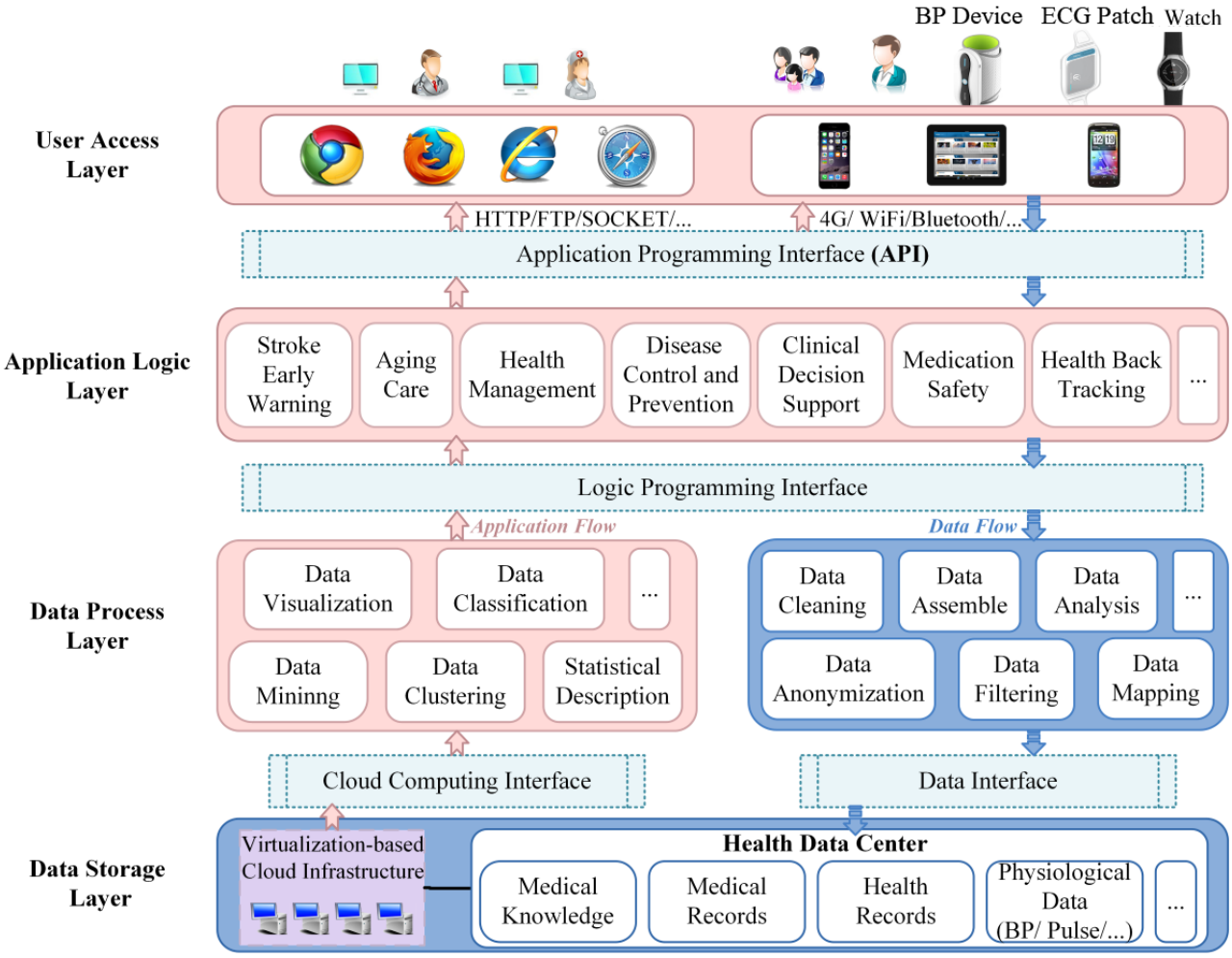
Ambulatory Blood Pressure Monitor Mobile Health System and Architecture

To enable the ABPM mHealth system to warn of early risk of stroke, a four-layer functional architecture was developed (Figure 1). These layers included the user access layer, the application logic layer, the data process layer and the data storage layer, the structure and function of which are described below.

User Access Layer

This layer provides the interfaces between the users/devices and the data center. In this layer, an ABPM device is used to collect data from potential stroke patients through online and application interfaces that are available to users, as shown in Figure 1. However, admission authority and transmission issues should be considered during this part of the process, as user password and text verification codes are used for identification, and Cross-Site Request Forgery defense and transport layer security are applied to secure information delivery.

Figure 1. The four-layer ambulatory blood pressure monitor mobile health architecture for stroke-risk early warning. BP: blood pressure; ECG: electrocardiogram.



Application Logic Layer

This layer involves multiple forms of business intelligence logic, such as health state maintenance, clinical decision support and treatment, and nursing and care. Among them, this paper focuses on an abnormal BP data analysis algorithm running in the data center for stroke early warning.

Data Process Layer

In this layer, all data are processed to remove electronic noise and motion interference. Then, the preprocessed signals are analyzed, mined, and visualized for the purpose of stroke early-risk warning.

Data Storage Layer

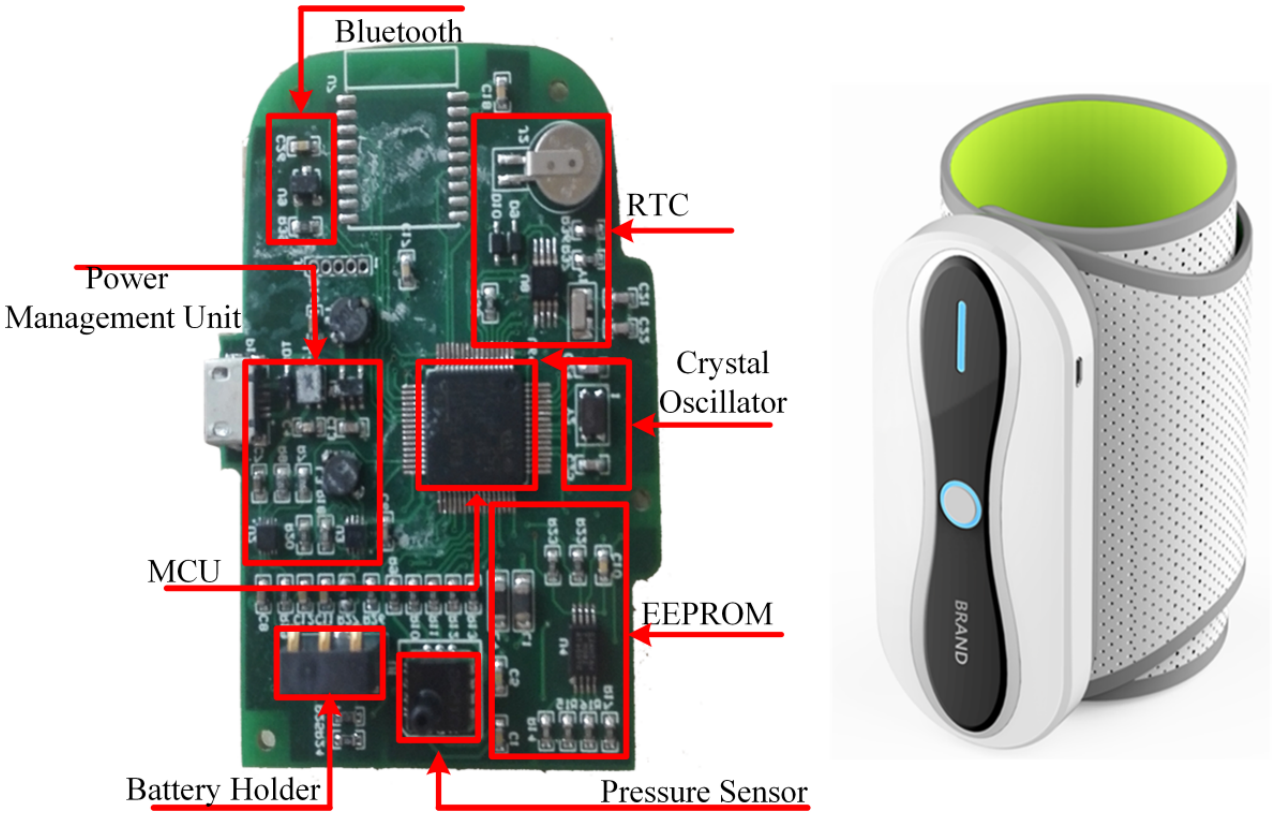
In this layer, it is necessary to handle missing or damaged data to guarantee data integrity.

In this study, a new ABPM mHealth architecture was developed to provide early stroke warning in real-time. Here, we focus on two major parts: at the frontier for real-time data collection, the device embedded with an inflation-type BP measurement algorithm, and at the data center, an abnormal BP data analysis algorithm for real-time early stroke warning using users' terminal interfaces.

At the Frontier: An Ambulatory Blood Pressure Monitor Device for Inflation-Type Blood Pressure Measurement

As previously mentioned, most pathological changes inside blood vessels may not be captured properly by BP devices with a deflation-type measurement. To solve this issue, our device is instead embedded with an inflation-type BP measurement. For different stroke patients, this device can be adapted to physiological and external factors as well. Additionally, our device can be automatically pumped to the appropriate maximum inflation pressure values according to the preset systolic blood pressure, which helps the new device adapt to users' health conditions for raw pressure data collection. For details about the inflation-type BP Measurement, please see Multimedia Appendix 1. Our developed ABPM device is shown in Figure 2.

Figure 2. Printed circuit board display of the new ambulatory blood pressure monitor device. RTC: real-time clock; MCU: microcontroller unit; EEPROM: electrically erasable programmable read-only memory.



There are multiple advantages to this device, including: (1) the adoption of the inflation-type measurement algorithm, which is the first commercial device that can capture vasculopathy information in detail; (2) that it is personalized and user-friendly, it can inflate to a suitable pressure level according to users' individual physiological states, and it automatically adjusts inflation speed according to users' activity states (6 mmHg/second while awake, and 4 mmHg/second during sleep), which can be recognized by a 3-axis sensor; and (3) the pump and cuff are bound together to reduce noise, in comparison with existing commercial products.

Early Warning Services for Stroke Risk at the Data Center

As mentioned above, stroke risk can be controlled if a Morning Blood Pressure Surge monitoring system can generate early warnings for stroke risk in real-time. To achieve this goal, we needed to detect abnormal stroke-related signals, which would then be uploaded to the data center. However, existing research for stroke prevention mainly focuses on statistical studies about public health. Kansadub et al investigated classification algorithms for a national- or group-level, stroke-risk increasing or reducing prediction model, which included methods such as naive Bayes, decision trees, and neural networks [28]. Yang et al constructed a predictive model for stroke based on multiple regression analyses of multiple climate factors [29]. Most of these studies focus on stroke risk factor analysis, not a real-time solution of abnormal stroke signal analysis.

In this study, an abnormal BP data analysis algorithm (see Multimedia Appendix 2) is used to provide early warning to patients. The algorithm classifies stroke risk levels as low, caution, or high. In this algorithm, a total of 88 features, including BP real-time data and health state information, are selected for classification, some of which are listed in Table 1 (health state information is from patient records). When this algorithm is running at the data center, all numerical values are normalized to 0 or 1 for empty numeric attributes, or they are normalized to the average value or 0 to replace empty Boolean attributes. This strategy is helpful for handling missing data in the algorithm. All necessary information is fed into the abnormal BP data analysis algorithm to determine the level of stroke risk. In this algorithm, the support vector machine approach is used to train data and simulated annealing is adopted to tune the hyper-parameters. Additionally, a 10-fold cross-validation was applied to refine the hyper-parameters.

Table 1. Input data and features (partial).

Feature, characteristic	Comments
24-hour ABPM[a]	
Hypertension grading	Calculated by WHO[b] hypertension definition
BP[c] load	Percentage of elevated pressure above a defined threshold
Dipper	Whether or not blood pressure falls at night compared to daytime values
Morning surge	Morning SBP[d]/preawake SBP
SBP, max	The highest SBP in a given time window
DBP[e], max	The highest DBP in a given time window
SBP, min	The lowest SBP in a given time window
DBP, min	The lowest DBP in a given time window
SBP, mean	Average SBP in a given time window
DBP, mean	Average DBP in a given time window
PP[f], mean	Average SBP minus average DBP
Personal information	
Age	Years
Gender	1=Man, 0=Woman
Height	Centimeters
Weight	Kilograms
Lifestyle	
Smoking	1=Yes, 0=No
Drink	1=Yes, 0=No
Physical inactivity	Amount of exercise per week
Patient history	
Stroke	1=Yes, 0=No
Diabetes mellitus	1=Yes, 0=No
High blood cholesterol	1=Yes, 0=No
Family history	
Stroke	Family members with stroke
Hypertension	Family members with hypertension

[a]ABPM: ambulatory blood pressure monitor.

[b]WHO: World Health Organization.

[c]BP: blood pressure.

[d]SBP: systolic blood pressure.

[e]DBP: diastolic blood pressure.

[f]PP: pulse pressure.

Design of Service Interfaces to Enable Stroke-Risk Alerts

In practice, stroke risk can be reduced if a Morning Blood Pressure Surge monitoring system can generate early risk warnings for stroke in real-time. At the data center, an application programming interface (API) was deployed to provide reliable and stable data transmission between users and the service center. We designed two types of user interfaces: one for doctors to manage and diagnose patients on the website, and the other for data presentation and analysis on users' phones (Figures 3 and 4).

Figure 3. The Web user interface of the Health Data Center.

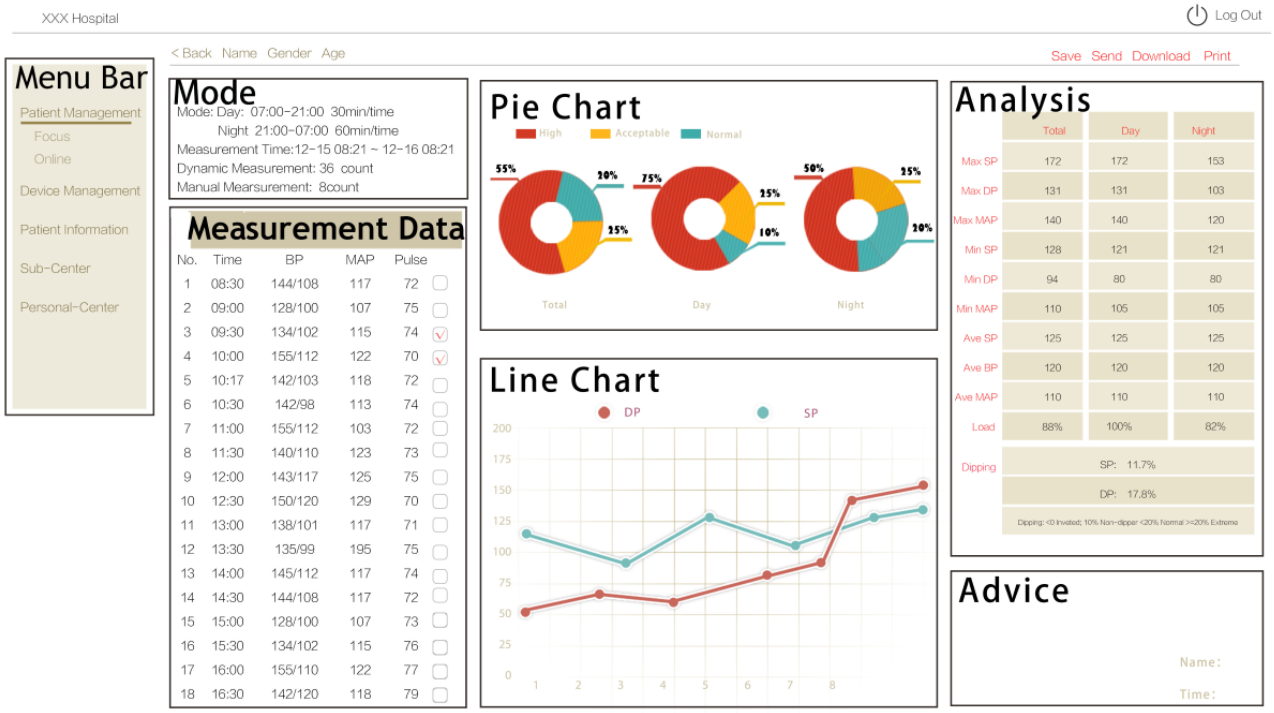
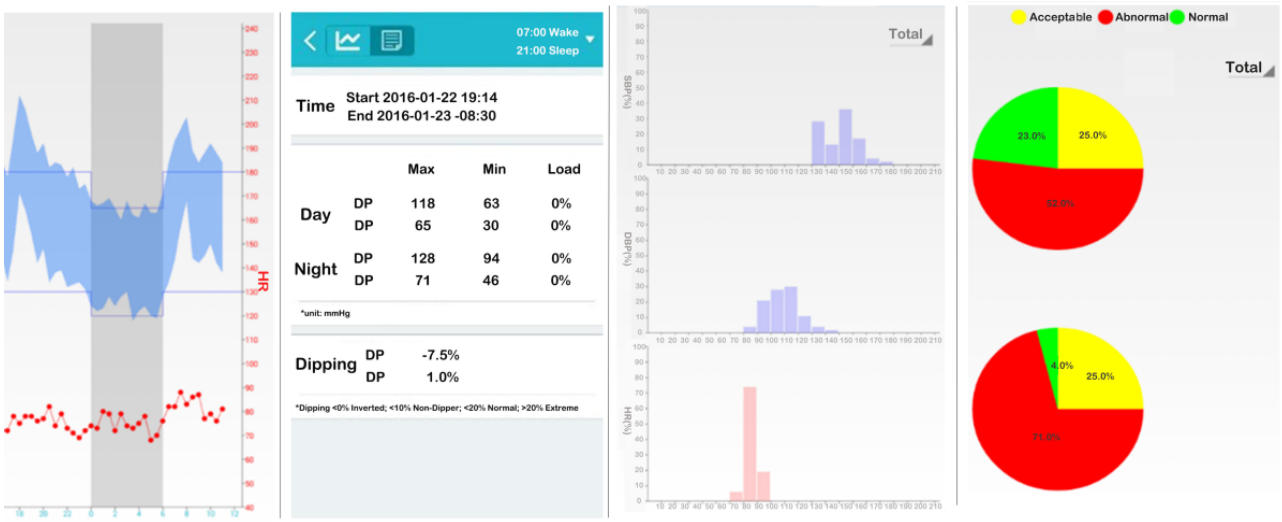


Figure 4. Data analysis and visualization interfaces for app users.



As shown in Figure 3, a doctor can review the medical signals, give instructions, or print a hard copy of diagnosis results. Users can display data in many forms on their phone, such as with histograms, pie charts, or line charts, as well as a 24-hour ABPM analysis report. As shown in Figure 4, the mobile app on smartphones provides patients with evaluation and visualization services, which are important for health management and stroke risk warning. This app has a robust network connection by using several different operations, including calling a local database instead of logging into the server again, providing rollback database operations to avoid inconsistency, using hash values to avoid repeated writing, etc.

Evaluation of the Ambulatory Blood Pressure Monitor Mobile Health System

Through a longitudinal observational study, our new inflation-based ABPM device and our stroke-risk early warning system were evaluated and compared against the gold standard mercury sphygmomanometer. All studies were carried out in the community hospitals of Peking University.

Firstly, to calibrate our device, we used the ProSim 8 simulator (FLUKE Corporation, Everett, Washington, United States) to generate a series of source-simulated signals (pulse/sleevelet waves) for calibrating the accuracy. In this study, the HEM-7207 electronic sphygmomanometer (Omron Corporation, Osaka, Japan) was used as a reference as it is in common use. The simulation data sets contain 20 variables: 7 standard BPs, 7 patient conditions, 3 arrhythmias, and 3 respiratory types related

to BP. The corresponding systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate ranges in the data set are 60-255 mmHg, 30-195 mmHg, and 40-180 times/minute, respectively (see Multimedia Appendix 1 for the results of our device tests).

Secondly, we further conducted device tests according to a gold standard for agreement analysis. According to the regulations from the China Food and Drug Administration which have been in place since 2015, a new BP device can be approved if tests are passed by using simulation signals [30]. This is the reason that simulation signals for device testing are enough to verify a BP device, without the need for real-world clinical trials. In this study we still performed 100 participant tests in Peking University community hospitals, and they were used for correlation coefficient and agreement analysis. A total of 120 participants were involved in these tests, including 56 females and 64 males. The mean age of participants was 49.5 years old (SD 15.6; range 22-86 years old), the mean weight was 68.1 kilograms (SD 19.7; range 45-103 kilograms), the mean height was 166 centimeters (SD 10.7; range 151-187 centimeters), and the average heart rate was 70.6 beats per minutes (SD 15.4; range 45-125 beats per minute). For each subject, BP was measured by the gold standard device, the mercury sphygmomanometer.

Finally, to validate the risk prediction function of this system, 209 patients were chosen as participants in Peking University community hospitals from April 12, 2017, to December 13, 2018. These subjects were required to use our app to upload personal health state information and wore the ABPM devices 24 hours a day to collect real-time BP data. Meanwhile, doctors made professional stroke-risk diagnoses of the subjects based on a comprehensive examination, including hematology, electrocardiograms, magnetic resonance angiograms, etc.

Results

Device Test Using Simulated Signals

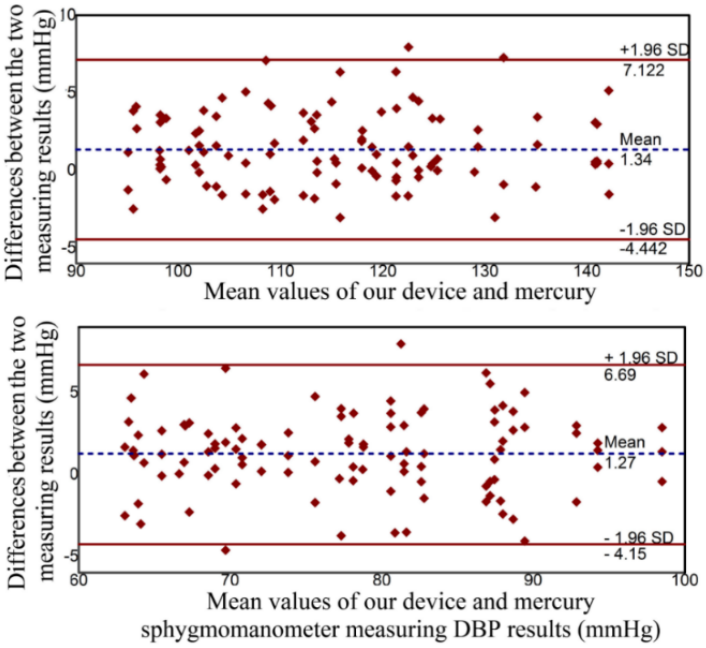
The device test using simulated signals showed that our new device outperformed the HEM-7207 electronic sphygmomanometer in most cases (for more detailed information see Multimedia Appendix 1). According to error analyses of the two devices, in our device the mean error for SBP was 0.55078 mmHg (SD 0.517412 mmHg) and for DBP it was 0.32185 mmHg (SD 0.39169 mmHg). However, the mean error for SBP and DBP in the HEM-7207 are 2.15 mmHg and 1.85 mmHg, respectively. It shows that our device has achieved measurement consistency and repeatability requirements.

Device Test Using a Standard Mercury Sphygmomanometer

We further conducted device tests for agreement analysis using the gold standard mercury sphygmomanometer. The measurement results showed that in our device the SBP was 125.45 mmHg (SD 25.2 mmHg) and the DBP was 77.92 mmHg (SD 23.5 mmHg), while for the mercury sphygmomanometer the SBP was 126.72 mmHg (SD 24.7 mmHg), and the DBP was 79 mmHg (SD 23.8 mmHg). The SBP difference between our device and the mercury sphygmomanometer was 1.34 (SD 2.95), while the DBP difference was 1.27 (SD 2.77). The correlation coefficient of our device with the mercury sphygmomanometer was 0.958 during the SBP test and 0.912 during the DBP test. This means that there is a high linear correlation between the two measurement approaches. In terms of the difference between the two measuring devices, their mean and standard deviation values are acceptable for practical applications.

As shown in Figure 5, the consistency and agreement of our device with mercury was evaluated. In this figure the Bland and Altman plot [31] was adopted, wherein the X-axis represents the mean of the SBP/DBP measurements of our device and the mercury sphygmomanometer and the Y-axis shows the difference between the two measurement approaches. The results of the correlation coefficient and agreement analysis clearly show that our new device met the clinical requirements and can be used to support clinical diagnosis.

Figure 5. Agreement analysis between our device and the sphygmomanometer using a Bland and Altman plot. DBP: diastolic blood pressure.



Longitudinal Observational Studies of the Ambulatory Blood Pressure Monitor System

We randomly selected 20% of subjects out of the 209 participants as a test set, another 20% as a validation set, and the remaining participants were used as a training set. To evaluate the performance of our risk prediction model, we compared our abnormal BP data analysis algorithm with fully connected neural networks and random forests. To evaluate the performance of our proposed methods and the alternatives, F1-score, specificity, accuracy, precision, recall, and area under the curve (AUC) parameters were adopted (Table 2). As shown in Table 2, the proposed abnormal BP data analysis method achieved better performance for the three risk levels. To be specific, for the high-risk level the accuracy was 0.885 and the AUC was 0.912.

Figure 6 presents the Receiver Operating Characteristic (ROC) curve of three risk levels: low, caution, and high. We can see the ROC curve is smooth, which means that there is no overfitting. Moreover, the abnormal BP data analysis algorithm is closer to the top left corner compared to the other two risk detection models, which means that our system achieved better prediction performance.

Table 2. Test performance of the abnormal blood pressure data analysis algorithm compared to other models.

Risk levels, models	F1-score	Specificity	Accuracy	Precision	Recall	AUC[a]
Low						
FCNN[b]	0.645	0.867	0.835	0.588	0.714	0.863
RF[c]	0.552	0.702	0.725	0.419	0.809	0.859
ABA[d]	0.659	0.867	0.840	0.596	0.738	0.904
Caution						
FCNN	0.771	0.628	0.710	0.790	0.753	0.731
RF	0.571	0.785	0.565	0.794	0.446	0.636
ABA	0.786	0.629	0.725	0.795	0.776	0.756
High						
FCNN	0.528	0.936	0.875	0.560	0.500	0.827
RF	0.519	0.848	0.83	0.434	0.646	0.894
ABA	0.673	0.953	0.885	0.618	0.714	0.912

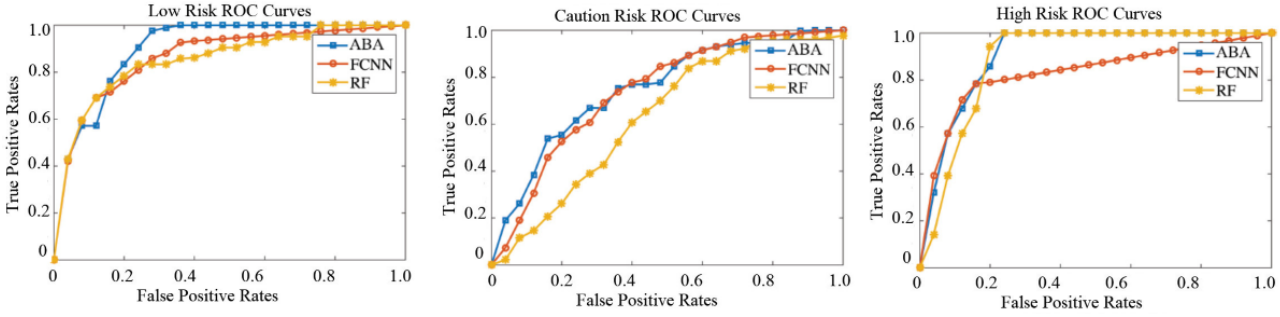
[a]AUC: area under the curve.

[b]FCNN: fully connected neural networks.

[c]RF: random forest.

[d]ABA: abnormal blood pressure data analysis.

Figure 6. ROC curves of stroke risk prediction. ROC: receiver operating characteristic; ABA: abnormal blood pressure data analysis; FCNN: fully connected neural networks; RF: random forest.



Discussion

Principal Results

With the application of mobile and communication technology in the medical field, mHealth, which can provide real-time health monitoring service for the elderly or patients with chronic diseases, has attracted a lot of research attention [32-35]. In this study, we proposed a novel ABPM mHealth system that can facilitate accurate BP measurement and generate early warnings for stroke risk in real-time. This paper describes the ABPM mHealth architecture, the device with its inflation-type BP measurement, and the stroke-risk prediction method at the data center. In the end, we carried out a longitudinal observational study and evaluated the proposed mHealth system.

Using mobile devices such as phones and patient monitoring devices, mHealth supports personal health management and public health activities [36]. Recent research has mainly focused on the technical aspect of research design, such as the Android platform, Internet of Things technology [37], cloud platforms [38], medical information systems, wireless monitoring equipment, Bluetooth technology, wearable devices [39], and so on. However, most of these studies are still in the pattern design stage or focus on how to develop or perfect the existing systems or mHealth applications, paying less attention to the user's demand.

We designed a novel ABPM mHealth system to carry out early warning of stroke risk. To capture and monitor a person’s BP profile and support clinical decision, a certified ABPM device (No. 20172201157) was designed and developed. Our device is customized, adaptive, and user-friendly, as it can pump according to a personal health condition, which allows for a personalized measurement with precision. After that, a risk prediction abnormal BP data analysis method is used to process the BP data, which enables this system to provide early warning of stroke risk to users in real-time.

As previously mentioned, a new inflation-type BP method was proposed in this study, and it was shown to be more effective compared to existing solutions. Currently there are two existing inflation-type BP methods, but neither is adaptive. The first one is the stop-and-wait inflation method [40] that contains several stopping points preset for an inflation process. During inflation, its pump is set to stop and wait for 2-3 seconds at each stopping point. In the 2-3 second waiting period, pressure sensor data is processed and then the algorithm decides whether to stop inflation. This method can reduce motion interference, but blood vessels under long-time compression increase inaccuracy. The second method is nonstop inflation [41] where the highest pulse wave peak is regarded as a criterion to calculate the mean arterial pressure (MAP), and then the MAP is used to determine the inflation-ending condition. Although the method is fast and simple, it is easily affected by motion interference since it introduces jitter signals that may be detected as the highest pulse wave peak. However, the detected highest pulse wave peaks in these cases do not represent a real criterion for MAP calculation. Therefore, the existing methods cannot be used to achieve accurate inflation-type BP measurement. To solve these existing problems, our work uses a sliding time window to optimize inflation where the cuff is linearly pumped by a stable pump algorithm and raw pressure data is collected and processed.

To evaluate the ABPM mHealth system performance, system evaluation and longitudinal observational studies were conducted, and 341 risk warning events were recorded, with these events then used to evaluate system performance in practice. Among these participants, there was a stroke patient who had 2-3 strokes per year over the past two years due to not following doctor instructions. This outpatient wore our device during the study, so their blood pressure was measured every day and medication ended up being taken whenever our system generated the early warning signals of a stroke. As a result, there were no recurrent stroke events for a whole year for this patient. Another important case was an emergency outpatient whose blood pressure was rising abnormally and triggered a warning alert in our system. The patient was sent to a hospital in time after our system notified his doctors and relatives. These cases show that our system can help patients realize their risk in real-time, which can reduce stroke risk and improve users' health conditions.

In addition to that, our system was able to generate periodic notifications and warnings about the patients' conditions. This ambulatory system could encourage users to pay attention to their medication states, diet, and emotional health, among other things, which could improve compliance with risk self-management. As a result, the new ABPM mHealth system equipped with the abnormal BP data analysis algorithm was proven to be a useful tool for controlling and preventing stroke, and it could be applied to detect early risks of cardiovascular disease and other BP-related diseases.

Limitations

This work is interdisciplinary in nature. Thus, we cannot cover all the details about the new ABPM mHealth system in one paper. This paper focused on how to construct an ABPM mHealth system for early warning of stroke risk. In terms of this system, the frontier (an inflation-type BP measurement device) and the data center (an abnormal BP data analysis algorithm for early risk alert) are highlighted in detail. The mHealth working principles are omitted because they follow standard practice right now.

To the best of our knowledge, this new ABPM device is the first commercial product that measures blood pressure during inflation. To improve measurement accuracy, a sliding window strategy that can improve inflation-type measurement and reduce motion interference substantially was adopted. Moreover, our ABPM device is equipped with a proportional integrative derivative controller responsible for linear inflation. The device has room for further improvement, such as insulation of mechanical noise, miniaturization in terms of size, reliability and stability of data transmission, battery consumption and continuation, multiple parameter processing, etc.

Most significantly, the abnormal BP data analysis algorithm can be used as clinical decision support, with a supervised machine learning algorithm to classify the risk level of a stroke patient in real-time. The algorithm has been implemented in our data center and was tested for performance. The performance of the abnormal BP data analysis algorithm was compared with the random forest algorithm and a fully connected neural network, showing that our abnormal BP data analysis algorithm has better performance in accuracy, precision, and recall. However, there remains work left for future studies, such as the development of an algorithm without the need for decision support from a remote data center, how to offer privacy protection, and so on.

In this study, we recruited 209 patients as longitudinal observational subjects in the community, which is not a big number in terms of clinical trials. This limited the training of a more accurate stroke early warning model and the validation of the performance of the ABPM mHealth system. We plan to carry out the study with a more comprehensive experimental design to test our ABPM mHealth system.

Acknowledgments

The work was supported by the National Natural Science Foundation of China Projects (#61906105).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The new device with an embedded inflation-type BP measurement algorithm: design and test.
[PDF File (Adobe PDF File), 572 KB - mhealth_v7i10e14926_app1.pdf]

Multimedia Appendix 2
An abnormal BP data analyzing algorithm for stroke-risk early warning.
[PDF File (Adobe PDF File), 239 KB - mhealth_v7i10e14926_app2.pdf]

References

1. Sims N, Muyderman H. Mitochondria, oxidative metabolism and cell death in stroke. Biochim Biophys Acta 2010 Jan;1802(1):80-91 [FREE Full text] [doi: 10.1016/j.bbadis.2009.09.003] [Medline: 19751827]
2. Chen Z, Jiang B, Ru X, Sun H, Sun D, Liu X, et al. Mortality of Stroke and Its Subtypes in China: Results from a Nationwide Population-Based Survey. Neuroepidemiology 2017;48(3-4):95-102. [doi: 10.1159/000477494] [Medline: 28586776]
3. World Health Organization. 2013 Apr. A global brief on hypertension: silent killer, global public health crisis URL: https://www.who.int/cardiovascular_diseases/publications/global_brief_hypertension/en/ [accessed 2019-10-22]
4. Khalid S, Soomro B, Mahmood S, Abbass A. Frequency of depression in patients with stroke. Pakistan Journal of Neurological Sciences 2015;10(2):13-17.
5. Zeng L, Meng C, Liang Z, Huang X, Li Z. Stroke unit of integrative medicine for post stroke comorbid anxiety and depression: A systematic review and meta-analysis of 25 randomized controlled trials. 2014 Presented at: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2-5 November; Belfast. [doi: 10.1109/BIBM.2014.6999360]
6. Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, et al. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol 2013 Dec;74(6):778-785 [FREE Full text] [doi: 10.1002/ana.23972] [Medline: 23868398]
7. Benjamin E, Muntner P, Alonso A, Bittencourt M, Callaway C, Carson A, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019

Update: A Report From the American Heart Association. Circulation 2019 Mar 05;139(10):e56-e528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
8. Tsivgoulis G, Pikilidou M, Katsanos A, Stamatelopoulos K, Michas F, Lykka A, et al. Association of Ambulatory Blood Pressure Monitoring parameters with the Framingham Stroke Risk Profile. J Neurol Sci 2017 Sep 15;380:106-111. [doi: 10.1016/j.jns.2017.07.007] [Medline: 28870547]
9. He Y, Yang M, Che S, Chen S, Jiang X, Guo Y. Effect of morning blood pressure peak on early progressive ischemic stroke: a prospective clinical study. Clin Neurol Neurosurg 2019 Sep;184:105420. [doi: 10.1016/j.clineuro.2019.105420] [Medline: 31310922]
10. Tsivgoulis G, Spengos K, Zakopoulos N, Manios E, Xinos K, Vassilopoulos D, et al. Twenty four hour pulse pressure predicts long term recurrence in acute stroke patients. J Neurol Neurosurg Psychiatry 2005 Oct;76(10):1360-1365 [FREE Full text] [doi: 10.1136/jnnp.2004.057265] [Medline: 16170077]
11. O'Donnell M, Xavier D, Liu L, Zhang H, Chin S, Rao-Melacini P, INTERSTROKE investigators. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet 2010 Jul 10;376(9735):112-123. [doi: 10.1016/S0140-6736(10)60834-3] [Medline: 20561675]
12. Argentino C, Toni D, Rasura M, Violi F, Sacchetti ML, Allegretta A, et al. Circadian variation in the frequency of ischemic stroke. Stroke 1990 Mar;21(3):387-389. [doi: 10.1161/01.str.21.3.387] [Medline: 2309262]
13. Manfredini R, Boari B, Portaluppi F. Morning surge in blood pressure as a predictor of silent and clinical cerebrovascular disease in elderly hypertensives. Circulation 2003 Sep 09;108(10):e72-e73. [doi: 10.1161/01.CIR.0000089096.13946.CB] [Medline: 12963687]
14. Huang A, Chen C, Bian K, Duan X, Chen M, Gao H, et al. WE-CARE: an intelligent mobile telecardiology system to enable mHealth applications. IEEE J Biomed Health Inform 2014 Mar;18(2):693-702. [doi: 10.1109/JBHI.2013.2279136] [Medline: 24608067]
15. Bondaronek P, Alkhaldi G, Slee A, Hamilton F, Murray E. Quality of Publicly Available Physical Activity Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
16. Pan D, Dhall R, Lieberman A, Petitti D. A mobile cloud-based Parkinson's disease assessment system for home-based monitoring. JMIR Mhealth Uhealth 2015 Mar 26;3(1):e29 [FREE Full text] [doi: 10.2196/mhealth.3956] [Medline: 25830687]
17. Liu C, Lin I, Lu J, Cai D. A Smartphone App for Improving Clinical Photography in Emergency Departments: Comparative Study. JMIR Mhealth Uhealth 2019 Jul 31;7(7):e14531 [FREE Full text] [doi: 10.2196/14531] [Medline: 31368444]
18. Altintas E, Takoh K, Ohno Y, Abe K, Akagawa T, Ariyama T, et al. Wearable and low-stress ambulatory blood pressure monitoring technology for hypertension diagnosis. 2015 Presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2015; Milano p. 4962-4965. [doi: 10.1109/EMBC.2015.7319505]
19. Liu Z, Shu M. Design and implementation of GPRS ambulatory blood pressure monitoring system based on cloud platform. International Symposium on Computers and Informatics 2015. [doi: 10.2991/isci-15.2015.224]
20. Tsoi K, Yip B, Au D, Kuo Y, Wong S, Woo J. Blood Pressure Monitoring on the Cloud System in Elderly Community Centres: A Data Capturing Platform for Application Research in Public Health. Blood Pressure Monitoring on the Cloud System in Elderly Community Centres; 2016 Presented at: 7th International Conference on Cloud Computing and Big Data (CCBD); 16-18 November; Sanya. [doi: 10.1109/CCBD.2016.068]
21. Wu N, Gong E, Wang B, Gu W, Ding N, Zhang Z, et al. A Smart and Multifaceted Mobile Health System for Delivering Evidence-Based Secondary Prevention of Stroke in Rural China: Design, Development, and Feasibility Study. JMIR Mhealth Uhealth 2019 Jul 19;7(7):e13503 [FREE Full text] [doi: 10.2196/13503] [Medline: 31325288]
22. Lin Z, Chang C, Chou N, Lin Y. Bluetooth Low Energy (BLE) based blood pressure monitoring system. 2014 Presented at: 2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG); 23-25 April; Taipei. [doi: 10.1109/IGBSG.2014.6835225]
23. Li W, Luo Y, Chang Y, Lin Y. A wireless blood pressure monitoring system for personal health management. 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology; 31 August-4 September; Buenos Aires p. 2196-2199. [doi: 10.1109/IEMBS.2010.5625944]
24. Jinling Z, Yue L, Jiabao W. Design of electronic blood pressure monitoring system based on mobile telemedicine system. 2013 Presented at: ICME International Conference on Complex Medical Engineering; 25-28 May; Beijing, China p. 145. [doi: 10.1109/ICCME.2013.6548228]
25. Rotariu C, Pasarica A, Costin H, Adochiei F, Ciobotariu R. Telemedicine system for remote blood pressure and heart rate monitoring. 2011 Presented at: E-Health and Bioengineering Conference (EHB); 24-26 November; Iasi, Romania p. 1.
26. Fabian V, Havlik J, Dvorak J, Kremen V, Sajgalik P, Bellamy V, et al. Differences in mean arterial pressure of young and elderly people measured by oscilometry during inflation and deflation of the arm cuff. Biomed Tech (Berl) 2016 Dec 01;61(6):611-621. [doi: 10.1515/bmt-2015-0098] [Medline: 26943590]
27. Zheng D, Pan F, Murray A. Effect of mechanical behaviour of the brachial artery on blood pressure measurement during both cuff inflation and cuff deflation. Blood Press Monit 2013 Oct;18(5):265-271. [doi: 10.1097/MBP.0b013e3283651d55] [Medline: 23924706]

28. Kansadub T, Thammaboosadee S, Kiattisin S, Jalayondeja C. Stroke risk prediction model based on demographic data. 2015 Presented at: 8th Biomedical Engineering International Conference (BMEiCON); 25-27 November; Pattaya, Thailand p. 1.
29. Yang J, Ji L, Wang Q, Lu X. The prediction model of stroke on climate factors by multiple regression. 2016 Presented at: IEEE Information Technology, Networking, Electronic and Automation Control Conference; 20-22 May; Chongqing, China. [doi: 10.1109/itnec.2016.7560428]
30. Notification of the China Food and Drug Administration on issuing the catalogue of category II medical devices exempted from clinical trials (No.12 of 2014). URL: http://samr.cfda.gov.cn/WS01/CL0087/105224.html [accessed 2014-08-21]
31. Francq BG, Govaerts B. How to regress and predict in a Bland-Altman plot? Review and contribution based on tolerance intervals and correlated-errors-in-variables models. Stat Med 2016 Jun 30;35(14):2328-2358. [doi: 10.1002/sim.6872] [Medline: 26822948]
32. Krishnamurti T, Davis AL, Wong-Parodi G, Fischhoff B, Sadovsky Y, Simhan HN. Development and Testing of the MyHealthyPregnancy App: A Behavioral Decision Research-Based Tool for Assessing and Communicating Pregnancy Risk. JMIR Mhealth Uhealth 2017 Apr 10;5(4):e42 [FREE Full text] [doi: 10.2196/mhealth.7036] [Medline: 28396302]
33. Stanton M, Molineux A, Mackenzie C, Kelly-Hope L. Mobile Technology for Empowering Health Workers in Underserved Communities: New Approaches to Facilitate the Elimination of Neglected Tropical Diseases. JMIR Public Health Surveill 2016;2(1):e2 [FREE Full text] [doi: 10.2196/publichealth.5064] [Medline: 27227155]
34. De Cock N, Vangeel J, Lachat C, Beullens K, Vervoort L, Goossens L, et al. Use of Fitness and Nutrition Apps: Associations With Body Mass Index, Snacking, and Drinking Habits in Adolescents. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e58 [FREE Full text] [doi: 10.2196/mhealth.6005] [Medline: 28442455]
35. Meng F, Guo X, Peng Z, Lai K, Zhao X. Investigating the Adoption of Mobile Health Services by Elderly Users: Trust Transfer Model and Survey Study. JMIR Mhealth Uhealth 2019 Jan 08;7(1):e12269 [FREE Full text] [doi: 10.2196/12269] [Medline: 30622092]
36. Kay M, Santos J, Takane M. mHealth: New horizons for health through mobile technologies. Geneva, Switzerland: World Health Organization; 2011:366-371.
37. Chai PR, Zhang H, Baugh CW, Jambaulikar GD, McCabe JC, Gorman JM, et al. Internet of Things Buttons for Real-Time Notifications in Hospital Operations: Proposal for Hospital Implementation. J Med Internet Res 2018 Aug 10;20(8):e251 [FREE Full text] [doi: 10.2196/jmir.9454] [Medline: 30097420]
38. Gao F, Thiebes S, Sunyaev A. Rethinking the Meaning of Cloud Computing for Health Care: A Taxonomic Perspective and Future Research Directions. J Med Internet Res 2018 Jul 11;20(7):e10041 [FREE Full text] [doi: 10.2196/10041] [Medline: 29997108]
39. Murakami H, Kawakami R, Nakae S, Yamada Y, Nakata Y, Ohkawara K, et al. Accuracy of 12 Wearable Devices for Estimating Physical Activity Energy Expenditure Using a Metabolic Chamber and the Doubly Labeled Water Method: Validation Study. JMIR Mhealth Uhealth 2019 Aug 02;7(8):e13938 [FREE Full text] [doi: 10.2196/13938] [Medline: 31376273]
40. Zhang X, Song X, Gu J, Ye J, Chen S. Research of Electronic Sphygmomanometer Intelligent Aeration Based on Pulse Wave Identification [in Chinese]. Zhongguo Yi Liao Qi Xie Za Zhi 2015 Mar;39(2):90-94. [Medline: 26204735]
41. Wei W, Bao-Ming P, Bao-Yue G, Sheng-Jin L. Design and Realization of Electronic Sphygmomanometer Based on Inflation Measurement. Computer Systems and Applications 2012 Feb [FREE Full text]

Abbreviations

ABPM: ambulatory blood pressure monitor
AUC: area under the curve
BP: blood pressure
DBP: diastolic blood pressure
MAP: mean arterial pressure
mHealth: mobile health
PP: pulse pressure
ROC: receiver operating characteristic
SBP: systolic blood pressure
WHO: World Health Organization

Edited by G Eysenbach; submitted 04.06.19; peer-reviewed by S Omboni; comments to author 04.07.19; revised version received 12.08.19; accepted 24.09.19; published 30.10.19.

Please cite as:
Wang G, Zhou S, Rezaei S, Liu X, Huang A
An Ambulatory Blood Pressure Monitor Mobile Health System for Early Warning for Stroke Risk: Longitudinal Observational Study
JMIR Mhealth Uhealth 2019;7(10):e14926
URL: http://mhealth.jmir.org/2019/10/e14926/
doi:10.2196/14926
PMID:31670694

Original Paper

Do Daily Fluctuations in Psychological and App-Related Variables Predict Engagement With an Alcohol Reduction App? A Series of N-Of-1 Studies

Olga Perski[1], PhD; Felix Naughton[2], PhD; Claire Garnett[1], PhD; Ann Blandford[3], PhD; Emma Beard[1], PhD; Robert West[1], PhD; Susan Michie[4], DPhil

[1]Department of Behavioural Science and Health, University College London, London, United Kingdom

[2]School of Health Sciences, University of East Anglia, Norwich, United Kingdom

[3]UCL Interaction Centre, University College London, London, United Kingdom

[4]Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Olga Perski, PhD
Department of Behavioural Science and Health
University College London
1-19 Torrington Place
London, WC1E 6BT
United Kingdom
Phone: 44 02076791258
Email: olga.perski@ucl.ac.uk

Abstract

Background: Previous studies have identified psychological and smartphone app–related predictors of engagement with alcohol reduction apps at a group level. However, strategies to promote engagement need to be effective at the individual level. Evidence as to whether group-level predictors of engagement are also predictive for individuals is lacking.

Objective: The aim of this study was to examine whether daily fluctuations in (1) the receipt of a reminder, (2) motivation to reduce alcohol, (3) perceived usefulness of the app, (4) alcohol consumption, and (5) perceived lack of time predicted within-person variability in the frequency and amount of engagement with an alcohol reduction app.

Methods: We conducted a series of observational *N*-of-1 studies. The predictor variables were measured twice daily for 28 days via ecological momentary assessments. The outcome variables were measured through automated recordings of the participants' app screen views. A total of nine London-based adults who drank alcohol excessively and were willing to set a reduction goal took part. Each participant's dataset was analyzed separately using generalized additive mixed models to derive incidence rate ratios (IRRs) for the within-person associations of the predictor and outcome variables. Debriefing interviews, analyzed using thematic analysis, were used to contextualize the findings.

Results: Predictors of the frequency and amount of engagement differed between individuals, and for the variables 'perceived usefulness of the app' and 'perceived lack of time', the direction of associations also differed between individuals. The most consistent predictors of within-person variability in the frequency of engagement were the receipt of a daily reminder (IRR=1.80-3.88; $P<.05$) and perceived usefulness of the app (IRR=0.82-1.42; $P<.05$). The most consistent predictors of within-person variability in the amount of engagement were motivation to reduce alcohol (IRR=1.67-3.45; $P<.05$) and perceived usefulness of the app (IRR=0.52-137.32; $P<.05$).

Conclusions: The utility of the selected psychological and app-related variables in predicting the frequency and amount of engagement with an alcohol reduction app differed at the individual level. This highlights that key within-person associations may be masked in group-level designs and suggests that different strategies to promote engagement may be required for different individuals.

(JMIR Mhealth Uhealth 2019;7(10):e14098) doi:10.2196/14098

KEYWORDS

apps; behavior change; excessive alcohol consumption; engagement; mHealth; n-of-1; time series analysis

Introduction

Background

Excessive alcohol consumption is a public health priority and is implicated in substantial costs to the economy through lost productivity, crime, and health care costs [1,2]. Digital interventions, including websites, smartphone apps, and wearable devices, can increase access to behavioral support, have a low incremental cost once developed, and reduce stigma associated with help seeking in person [3-5]. Alcohol reduction apps have the added advantage of being available to users as and when needed. Some form of engagement, comprising both behavioral (eg, amount, depth, and frequency of app use) and experiential (eg, attention and interest) dimensions [6], is logically necessary for alcohol reduction apps to be effective [7,8]. Findings from an integrative literature review, in-depth interviews with potential users, theorizing within an interdisciplinary research team, and the development and evaluation of a novel self-report measure suggest that engagement with digital interventions can be defined as " a state-like construct which occurs each time a user interacts with a digital behavior change intervention with two behavioral (i.e. amount and depth of use) and three experiential (i.e. attention, interest and enjoyment) dimensions " (Perski et al, in press).

As observed levels of engagement with many digital interventions are considered too limited to support behavior change [9], efforts have been made to identify factors that predict engagement. Whether or not a user engages with a given digital intervention is likely to depend on its content (eg, behavior change techniques), how that content is delivered (eg, design features), the context in which the intervention is used (eg, who the users are and where they are using the intervention), whether or not the intervention succeeds in changing particular 'mechanisms of action' that mediate behavior change (eg, motivation and self-regulatory skills), and successful or unsuccessful behavior change (eg, the extent of alcohol reduction) [6]. To the authors' knowledge, studies to date have typically focused on the identification of group-level predictors of engagement with digital interventions for alcohol reduction [6]. As strategies to increase engagement need to be effective for individuals [10,11], it is important to examine whether key predictors identified at the group level are also predictive at the individual level.

Published secondary analyses of data from randomized controlled trials (RCTs) of digital interventions for alcohol reduction have identified group-level predictors of engagement. These studies show that demographic (eg, being female, older, and more highly educated) [12-14], psychological (eg, higher levels of baseline motivation to change) [13,15], drinking (eg, lower baseline levels of alcohol consumption), [12,13,16] and app-related variables (eg, the receipt of proactive reminders) [17] predict the total frequency and amount of engagement.

Qualitative studies asking excessive drinkers to reflect on the factors they expect to be the most important for engagement with apps for alcohol reduction have identified the following: motivation to change, perceived personal relevance of the app (defined as the extent to which the user believes that the app is suited to their individual needs [18]), and perceived usefulness of the app (defined as the extent to which the individual believes that use of the app will help them achieve their goal(s) [19,20]). Although common themes were pulled out from these qualitative studies, agreement among potential users on what factors are expected to be most important for engagement was low [20]. Qualitative research has also been conducted with participants who disengaged before the completion of an RCT of a Web-based alcohol reduction intervention [13]. When retrospectively asked to reflect on why they disengaged from the intervention, users frequently mentioned perceived lack of time (eg, being too busy and having other priorities), dissatisfaction with the intervention (eg, poor usability and irrelevant content), and improvement in the condition (eg, feeling better).

As mentioned, quantitative studies examining predictors of engagement have typically relied on group-level designs, aggregating data across participants. However, individual-level interventions, including alcohol reduction apps, are designed to target within-person processes that lead to behavior change. Intervention strategies aimed at increasing engagement (eg, proactive reminders, rewards, and feedback) need to be effective for individuals. It is, therefore, important to examine whether associations identified at the group level are also identified at the individual level. The *N*-of-1 study design, also known as a single-case design, is ideally suited for the assessment of within-person processes. The *N*-of-1 design can be either observational or experimental and "...receives its name by virtue of its sample size: *N* is equal to one" [21].

Previous qualitative and quantitative research has relied on either prospective or retrospective (as opposed to real-time) self-reports of psychological processes; these are likely to be biased or inaccurate [22]. For example, when prospectively predicting what factors are expected to be most important for engagement, potential users tend to highlight app-related aspects, such as the presence of features that enhance motivation to change (eg, goal setting, self-monitoring, and proactive reminders) and perceived usefulness (eg, tailoring of content and rewards) [18,20]. However, when asked to retrospectively report on factors they thought contributed to their disengagement from a digital intervention, different aspects tended to be highlighted, such as perceived lack of time [13]. A data gathering method that overcomes the problems associated with both prospective and retrospective self-reports is ecological momentary assessment (EMA), which involves the repeated measurement of psychological processes in real time [23,24]. Methods for the statistical analysis of data from EMA and *N*-of-1 studies include correlational and time series analyses [25,26], with the latter being an underused approach to date.

Objectives

This study used a series of *N*-of-1 studies, harnessing twice-daily EMAs for 28 days, and applied an innovative type of time series analysis to examine whether daily fluctuations in (1) the receipt of a reminder, (2) motivation to reduce alcohol, (3) perceived usefulness of the app, (4) alcohol consumption, and (5) perceived lack of time predicted within-person variability in the frequency (ie, number of log-ins) and amount (ie, time spent

per log-in) of engagement with a theory- and evidence-informed alcohol reduction app, *Drink Less* [27,28]. This study aimed to provide a greater understanding of the temporal direction of the relationships under investigation by assessing predictor variables before the measurement of outcome variables.

Methods

Study Design

A prespecified study protocol and analysis plan can be found on the Open Science Framework [29]. A series of observational *N*-of-1 studies was conducted with twice-daily (ie, morning and evening) assessments of psychological and app-related predictor variables. The outcome variables were the objectively estimated frequency and amount of engagement with the *Drink Less* app, described in detail in the *Measures* section below. Although the subjective experience (eg, attention and interest) is also thought to be a key dimension of digital engagement ([6]; Perski et al, in press), only behavioral indicators of engagement (which can be measured automatically via participants' app screen views) were considered in this study to minimize participant burden. Although it had been prespecified in the study protocol that the key outcome of interest was the frequency of engagement, a series of unplanned analyses with the variable 'amount of engagement' was also conducted. To help contextualize the quantitative findings, semistructured debriefing interviews were conducted over the phone after the 28-day study period.

Participants and Sampling

The eligibility criteria are outlined in Textbox 1. Participants were excluded if they were not fluent English speakers. Recruitment was conducted on the Web via the research platform *Call for Participants*, social media (ie, Twitter), and an alcohol reduction charity's mailing list. The recruitment materials stated that regular drinkers were invited to take part in a study on how people use alcohol reduction apps in their daily lives, which involved responding to twice-daily text messages for 28 days.

Textbox 1. Participant eligibility criteria.

Eligibility criteria:

- Aged 18 years and older
- Owned an Apple iPhone capable of running iOS v.8.0 software or higher (ie, iPhone 4S or later models)
- Resided in or near London and willing to come to University College London for a briefing interview (to ensure adequate study commitment)
- Reported an Alcohol Use Disorders Identification Test score of ≥8, indicating excessive alcohol consumption [30]
- Was interested in using an app to reduce their drinking
- Was willing to set a goal to reduce their drinking
- Installed the *Drink Less* app and opened it at least once following the briefing interview
- Was willing to engage with the app daily for 28 days, recognizing that there may be occasional days where they would not engage with it [31]
- Was willing to respond to twice-daily text messages for 28 days
- Was willing to take part in a debriefing interview conducted over the phone

The number of observations (and not the number of participants) determines the statistical power in *N*-of-1 studies [32]. Each participant was asked to respond to twice-daily EMAs for 28 days, resulting in up to 56 data inputs per participant. The measurement frequency of 2 EMAs per day was informed by previous research conducted within the behavioral science domain [33]. The study duration of 28 days was selected as this is a common duration for digital alcohol reduction interventions [34]. As data were planned to be analyzed using generalized additive mixed models (GAMMs; see the *Data Analysis* section below), Monte Carlo simulations [35] estimated the statistical power achieved with a total of 56 data inputs. The power analysis, conducted in R, indicated that the study would have 80% power to detect an incidence rate ratio (IRR) of 1.8 for the association between 'perceived usefulness of the app' (predictor variable) and 'frequency of engagement' (outcome variable). Given the uncertainties regarding the distribution of model parameters, this power analysis should be interpreted with caution. See Table 1 for details about statistical assumptions used to inform the power analysis. To allow for a descriptive (but not inferential) comparison of potential between-person differences in the associations between the predictor variables and app engagement, a total of 8 participants was considered sufficient. As previous *N*-of-1 studies report up to 47% study dropout [33,36,37], we aimed to recruit an additional 50% of the target sample (ie, 12 participants).

Table 1. Statistical assumptions used to inform the simulation-based power analysis.

Considerations	Statistical assumptions and source of information (where available)
Model type	Generalized additive mixed model
Number of observations	Twice-daily ecological momentary assessments for a period of 28 days (ie, a total of 56 data inputs per participant)
Seasonality	No seasonality reflected by the day of the week the data were collected.
Distribution and point estimate (outcome variable)	The outcome variable (ie, frequency of engagement, operationalized as the number of app log-ins per measurement period) was assumed to follow a Poisson distribution with a mean of 11.7 log-ins per measurement period [28]. As the outcome variable represented count data, it was expected to follow a Poisson distribution. The mean of 11.7 log-ins was drawn from a group-level, factorial screening experiment of the *Drink Less* app [28], as this was judged to represent the best available data.
Distribution and point estimate (predictor variable)	The predictor variable (ie, perceived usefulness of the app), selected as a basis for the power analysis as data on the relationship of the other predictors and the frequency of engagement were lacking in the extant literature, was assumed to follow an autoregressive (AR) integrated moving average (MA) process with first-order autocorrelation, as it was expected that measurements would be similar to those taken 12 hours previously. We drew on the results from the between-person, factorial screening experiment of the *Drink Less* app, which assessed the variable 'helpfulness of the app' at 28-day follow-up. This variable was deemed to be conceptually similar to the target variable. It was, therefore, assumed that the mean level of the predictor variable would be 3.18 (SD 0.93) [28].

Intervention

The *Drink Less* app is a stand-alone intervention designed to promote alcohol reduction in adults who drink excessively. The app is centered around a goal-setting module that allows users to select 1 or multiple weekly goals of their choice (eg, maximum number of units, alcohol-free days, spending on alcohol, or number of alcohol-attributed calories). The app includes 5 additional intervention modules: (1) normative feedback (ie, a visual gauge of how users' drinking compares with that of others in the same gender and age group), (2) cognitive bias retraining (ie, a game that aims to help users retrain automatic approach/attentional biases toward alcohol-related cues), (3) self-monitoring and feedback (ie, an interactive calendar that allows users to record and visualize drinks consumed/alcohol-free days), (4) action planning (ie, a feature that explains the benefits of setting if-then rules and allows users to create, review, and edit these), and (5) identity change (ie, a feature that allows users to view pairs of positive and negative outcome expectancies, record video messages to watch at a later date, and identify and select values of importance to their identity). Details about how intervention content was selected [38,39], user feedback on an early version of the app [40], the development process [27], and the first evaluation of the app's components in a randomized, factorial screening experiment [28] have been described in detail elsewhere. The *Drink Less* app allows users to set a daily reminder to open the app, which can be switched on or off and set to a suitable timing.

Measures

The following data were collected at baseline to determine study eligibility and to describe the sample: (1) age, (2) gender, (3) type of work (ie, manual, nonmanual, or other), (4) whether the participants owned an iPhone capable of running iOS 8.0 software or higher (ie, iPhone 4S or later models), (5) whether the participants were residing in or near London and were willing to come to University College London (UCL) for a briefing interview, (6) alcohol consumption, measured using the Alcohol Use Disorders Identification Test [30], a 10-item measure of alcohol consumption, drinking behavior, and alcohol-related problems, which provides a score ranging from 0 to 40, with scores ≥8 indicating excessive alcohol consumption, (7) whether the participants were interested in using an app to reduce their drinking, (8) whether the participants were willing to set a goal to reduce their drinking, (9) whether the participants were willing to engage with the study app daily for 28 days, (10) whether the participants had previously used an alcohol reduction app and if so, which one, (11) whether the participants were willing to respond to the twice-daily text messages for 28 days, and (12) whether the participants were willing to take part in a poststudy interview conducted over the phone.

Ecological Momentary Assessments (Predictor Variables)

The following data were collected twice per day (ie, morning and evening):

- Motivation to reduce alcohol was measured by asking "How motivated are you currently to reduce your drinking?" The response options ranged from 1 to 7, with 1 indicating *not at all* and 7 indicating *extremely*.
- Perceived usefulness of the app was measured by asking "How useful do you currently think the *Drink Less* app is for you?" The response options ranged from 1 to 7, with 1 indicating *not at all* and 7 indicating *extremely*. The decision to focus on perceived usefulness of the app in this study was informed by a meta-analysis of 59 studies indicating that the variable 'perceived usefulness' is consistently associated with behavioral intentions to use technology (r=.59) [41]; less is known about the relationship between the variable 'perceived relevance' and key outcome variables. This variable captured participants' beliefs about the app's usefulness and was considered in the absence of any objective effectiveness data from a confirmatory RCT.
- Alcohol consumption was measured by asking "How many drinks containing alcohol have you had in the past 12 hours?" Participants were instructed to input integers only (ie, whole drinks).

- Perceived lack of time was measured by asking participants "To what extent do you currently have time for the *Drink Less* app?" The response options ranged from 1 to 7, with 1 indicating *I do not have any time for the app* and 7 indicating *I have lots of time for the app*.

An additional predictor variable, tailored to participants' preferences, was as follows:

- Whether or not a proactive reminder was received during each 12-hour measurement period; this variable was coded 1 if a reminder was received and 0 if it was not received. A maximum of 1 reminder could be received every 24 hours, and the frequency and the timing of the reminders did not change during the course of the study.

Outcome Variables

App screen views were automatically recorded, stored in a Web-based database, and extracted using the free python library *pandas* to derive the outcome variables frequency of engagement and amount of engagement. The variable 'frequency of engagement' was operationalized as the number of log-ins during each 12-hour measurement period, with a log-in defined as a new screen view following at least 30 min of inactivity [42]. The variable 'amount of engagement' was derived by calculating the time spent (in seconds) per 12-hour measurement period. For descriptive purposes, the variable 'depth of engagement' was also derived, which was operationalized as the number of app components accessed per 12-hour measurement period, indexed as a proportion of the number of available app components. However, as depth of engagement was strongly correlated with amount of engagement for all participants, no inferential analyses were conducted using this variable.

Procedure

Participants who expressed an interest in taking part were asked to read the participant information sheet, provide informed consent, and fill out the Web-based screening questionnaire hosted via Qualtrics [43]. Eligible participants were invited to a briefing interview at UCL where they were asked to reread the information sheet and were consented. Participants were asked to download the *Drink Less* app, briefly explore it, and set at least 1 weekly alcohol reduction goal of their choice. They were also asked if they wanted to switch the daily reminder on or off and if applicable, select a suitable timing for these. After having explored the app, participants were asked to complete a brief survey on their phone, which fetched their unique user identity document, generated by the *Drink Less* app. This information enabled the researchers to match participants to their app screen views and, hence, derive the outcome variables. Participants were asked a few questions about their expected app use and what they were hoping to achieve using the app (not reported). They were subsequently asked to familiarize themselves with the daily EMA questions and response options and practice inputting their responses to the 4 questions into a single text message. They were also asked to select a suitable timing for the EMAs. In the morning, participants were asked to select a time between 6 am and 10 am and in the evening, between 6 pm and 10 pm, ensuring that the selected time points did not fall earlier/later than their usual morning and evening bedtimes, respectively. No particular instructions about app engagement were provided other than that participants were expected to engage with the app at least once daily for 28 days, recognizing that there might be occasional days when they would not engage with it. Participants were told that they had to respond to at least 70% of the text messages and take part in the debriefing interview to receive any payment. They were also asked to notify the study team if they decided to change the timing of the daily reminder, so that this could be accounted for in the statistical analyses. The briefing interviews lasted between 29 and 63 min.

Participants were then asked to respond to the twice-daily text messages for 28 days, sent manually from an iPhone 6S by the first author. The first text message was sent the morning after the briefing interview. When a response was received, participants were sent the following standard response: "Thank you for your responses!" Participants also received weekly updates via text message about their survey response rate to encourage adherence to the study materials (eg, "Hi X! Thank you for completing the first week of the study. You have responded to X/14 text messages. Keep up the good work!"). If the text messages were not received in the expected format, participants received a standard reply with instructions for how to input the responses (ie, "Hi X! It appears that your responses are not in the expected format. Please enter your responses as follows: a=X; b=X; c=X; d=X").

After 28 days, participants were invited to take part in a debriefing interview conducted over the phone, during which they were asked about their experiences of engaging with the *Drink Less* app. The interviews lasted between 25 and 47 min.

Participants were paid £0.50 per data input (ie, a maximum of £28), in addition to £32 upon study completion, resulting in a possible total of £60. This was paid to participants in the form of a shopping voucher.

Data Analysis

Guided by published research in the behavioral science domain [33,36,37], in time series with >5% missing data, multiple imputation was conducted using an expectation-maximization with bootstrapping algorithm via the R package *Amelia II*. Data were imputed separately for each dataset (ie, each participant). A polynomial time trend (ie, linear or quadratic) was included if this was found to improve the precision of the imputed data points. This was decided upon by examining the 95% CIs of the means of the imputed data points. A total of 5 imputed datasets were created per dataset with missing values, which were combined before conducting further statistical analyses using Rubin rules [33,36,37].

Descriptive statistics were calculated for each participant. Time series analyses were conducted using the R package *mgcv*: GAMMs were fitted to estimate IRRs for the associations between the predictor and the outcome variables. The IRR is a measure of relative difference and can, in this particular context, be interpreted as the relative frequency or amount of engagement for the different levels of the predictor variables. The GAMM is a type of multilevel model that has previously been applied to data from *N*-of-1 studies [44]. GAMMs are particularly well

suited to the modeling of time series data with 1 level of measurement (ie, repeated measurements nested within 1 individual), as they accommodate the inclusion of autocorrelation [44]. The analyses proceeded in a number of stages using a backward selection procedure:

- As the outcome variables represented counts, data were first assessed for overdispersion (ie, when the variance is greater than the mean). If there was evidence for overdispersion, a quasi-Poisson distribution (rather than a Poisson distribution) was specified.
- As repeated measures taken from the same individual are often correlated, data from *N*-of-1 studies typically violate the assumption of independence of observations. Autocorrelation was therefore assessed through the autocorrelation function and the partial autocorrelation function. Evidence of first-order autocorrelation in this study would mean that measurements were significantly correlated with those taken 12 hours previously.
- A full model including all predictor variables was first fitted to determine the most appropriate autocorrelation structure for each participant. Model fit was compared using Akaike's Information Criterion [45]. Although the *a priori* power analysis did not take account of the adjustment for seasonality or moving average (MA) terms, it was determined *a posteriori* that adjusting for the day of the week through the inclusion of a cyclic cubic smoothing term significantly improved the model fit for all participants and that the inclusion of an MA term improved the model fit for some participants.
- For visualization purposes, univariable models for each predictor variable were fitted for each participant, carrying forward the most appropriate autocorrelation structure and MA terms from the previous step.
- Parsimonious multivariable models were subsequently built through the stepwise elimination of redundant terms. The predictor variables were sequentially varied to arrive at a best-fitting model for each participant.

Debriefing Interviews

Telephone interviews were audio-recorded, transcribed verbatim by the first author, and analyzed using inductive thematic analysis [46], which involved the following steps: (1) data familiarization, (2) initial code generation, (3) searching for themes, (4) reviewing the themes, (5) defining and naming the themes, and (6) producing the report. Data were coded by the first author and reviewed by the third author. New inductive codes were labeled as they were identified during the coding process. Codes were subsequently reviewed one by one and systematically organized into themes.

Ethical Approval

Ethical approval was granted by UCL's Computer Science Departmental Research Ethics Chair (Project ID: UCLIC/1617/004/Staff Blandford HFDH). Personal identifiers were removed and anonymized data were stored securely on a password-protected computer. Participants' contact details were stored separately in a locked cabinet. The subscriber identification module card used to deliver the daily text messages was wiped upon completion of the data collection.

Results

Participants

Of the 22 participants who completed the Web-based screening questionnaire, 11 met the inclusion criteria and were invited to take part. Of these, 1 was unable to initiate the 28-day study during the planned study period. In total, 10 participants took part between June 29 and August 9, 2018. A participant broke their phone 14 days into the study and redownloaded the app onto a new phone without notifying the researchers. Owing to these technical issues, the new phone's app screens failed to sync with the database, and hence, the outcome data for the last 14 days of the study were lost. This participant was therefore excluded from the inferential analyses, but descriptive statistics were calculated for all 10 participants. Participants' characteristics are summarized in Table 2.

Table 2. Participants' demographic, drinking, and app-related characteristics.

Participant (P) identifier	Gender	Age (years)	Occupational status	Alcohol Use Disorders Identification Test	Past use of an alcohol reduction app	Past use of the *Drink Less* app
P1	Female	28	Nonmanual	16	No	No
P2	Female	20	Other	10	No	No
P3	Female	25	Nonmanual	30	No	No
P4	Female	18	Other	12	No	No
P5	Male	21	Other	22	No	No
P6	Female	31	Nonmanual	8	No	No
P7	Female	23	Nonmanual	12	Yes	Yes
P8	Female	30	Nonmanual	11	No	No
P9	Female	28	Other	23	Yes	No
P10	Female	26	Nonmanual	10	No	No

Descriptive Statistics

A total of 8 participants (8/10, 80%) opted to have the daily reminder switched on. Overall, participants displayed high compliance with the daily text messages (mean 93%; SD 5.8%), with the number of missing responses varying from 0% to 16% (see Table 3). Descriptive statistics for the predictor variables are displayed in Table 4.

Participants' total number of log-ins ranged from 10 to 69 (see Table 5). The total depth of engagement over the 28-day study period ranged from 14% (ie, accessing 1 of the app's 7 components) to 86% (ie, accessing 6 of the app's 7 components), and the total amount of engagement ranged from 4 min and 24 seconds to 70 min and 14 seconds. See Multimedia Appendix 1 for plots of participants' frequency and amount of engagement over the course of the study.

Table 3. Compliance with the twice-daily ecological momentary assessments.

Participant (P) identifier	Compliance (N=56), n (%)	Timing of text messages	Daily reminder switched on/off	Timing of daily reminder
P1	56 (100)	10 am/pm	On	10 am
P2	55 (98)	10 am/pm	On	1 pm
P3	50 (89)	7:30 am/pm	On	4 pm
P4	49 (88)	10 am/pm	On	11 am
P5	55 (98)	9:30 am/pm	Off	—[a]
P6	47 (84)	10 am/pm	On	10 am
P7	48 (86)	9 am/pm	On	9 am
P8	51 (91)	10 am/pm	Off	—
P9	56 (100)	10 am/pm	On	10:30 am
P10	54 (96)	10 am/pm	On	9 am

[a]Not applicable.

Table 4. Descriptive statistics for the predictor variables.

Participant (P) identifier	Motivation to reduce alcohol		Perceived usefulness of the app		Alcohol consumption (drinks)		Perceived lack of time	
	Mean[a] (SD)	Range	Mean[a] (SD)	Range	Mean[a] (SD)	Range	Mean[a] (SD)	Range
P1	5.3 (1.1)	3-7	5.4 (0.8)	4-7	2.1 (2.8)	0-10	6.1 (1.2)	3-7
P2	6.3 (1.1)[b]	3-7	6.3 (1.1)[b]	3-7	0.1 (0.5)[b]	0-3	4.6 (2.2)[b]	1-7
P3	5.2 (0.9)[b]	4-7	5.3 (1.1)[b]	3-7	1.2 (1.3)[b]	0-5	4.5 (1.0)[b]	2-7
P4	4.1 (1.6)[b]	1-7	2.4 (1.3)[b]	1-5	0.1 (0.8)[b]	0-4	4.9 (1.8)[b]	2-7
P5	3.6 (1.0)[b]	2-6	3.6 (1.2)[b]	1-7	1.2 (1.7)[b]	0-8	3.9 (0.9)[b]	2-7
P6	5.6 (0.7)[b]	4-7	4.4 (0.6)[b]	4-6	0.3 (0.8)[b]	0-3	4.4 (0.7)[b]	3-7
P7	4.1 (1.2)[b]	1-6	3.2 (0.9)[b]	2-5	1.1 (2.1)[b]	0-6	2.8 (1.6)[b]	1-6
P8	5.9 (0.5)[b]	4-7	6.1 (0.9)[b]	4-7	0.4 (0.9)[b]	0-4	2.2 (1.4)[b]	1-5
P9	4.3 (1.9)	1-7	1.9 (0.9)	1-5	3.9 (4.3)	0-14	6.0 (1.3)	2-7
P10	5.3 (1.6)[b]	1-7	4.8 (1.0)[b]	1-6	1.9 (2.9)[b]	0-9	5.5 (1.0)[b]	3-7

[a]Mean levels for the predictor variables over the 56 12-hour measurement periods.

[b]For participants with missing data, means and standard deviations for the complete datasets (after multiple imputation) were computed using Rubin rules.

Table 5. Descriptive statistics of participants' frequency, amount, and depth of engagement with the *Drink Less* app.

Participant (P) identifier	Log-ins over the 28-day study			Total amount of engagement over the 28-day study (minutes:seconds)	Amount of engagement per measurement period (minutes:seconds)		Total depth of engagement over the 28-day study (%)	Depth of engagement per measurement period (%), mean (SD)
	Total number	Mean (SD)	Range		Mean (SD)	Range		
P1	39	0.7 (0.7)	0-3	23:11	00:26 (00:53)	00:00-04:12	71	10 (12)
P2	47	0.8 (0.8)	0-4	60:43	01:06 (02:33)	00:00-16:32	86	20 (20)
P3	35	0.6 (0.6)	0-2	13:12	00:14 (00:27)	00:00-02:19	57	10 (11)
P4	10	0.2 (0.5)	0-2	04:24	00:05 (00:18)	00:00-01:29	43	3 (8)
P5	42	0.8 (0.7)	0-3	18:20	00:20 (00:29)	00:00-01:11	29	11 (11)
P6	31	0.6 (0.6)	0-2	39:19	00:42 (85.42)	00:00-08:12	57	9 (11)
P7	64	1.1 (0.9)	0-3	19:14	00:21 (00:27)	00:00-02:44	14	10 (6)
P8	69	1.2 (0.9)	0-3	70:14	01:09 (02:01)	00:00-10:47	43	17 (13)
P9	34	0.6 (0.7)	0-2	35:26	00:38 (02:04)	00:00-13:40	43	9 (11)
P10	__[a]	—	—	—	—	—	—	—

[a]Due to a technical issue, data were lost for P10.

Predicting the Frequency and the Amount of Engagement

The results from the univariable GAMMs can be found in Multimedia Appendix 2. For visualization purposes, plots of the IRRs and 95% CIs are depicted in the below figures. Table 6 reports the results from the multivariable GAMMs. In some cases, results from the univariable and multivariable models differed. Hence, interpretations are based on both uni- and multivariable analyses.

Table 6. Incidence rate ratios for the associations between the predictor and the outcome variables for each participant (P) in the multivariable generalized additive mixed models.

Participant	Frequency of engagement[a]		Amount of engagement[a]	
	Incidence rate ratio (IRR) (95% CI)	*P* value	IRR (95% CI)	*P* value
P1				
Reminder	$1.80_{2,1}$[b] (1.19-2.74)	*.01*[c]	—[d]	—
Motivation to reduce alcohol	$1.14_{2,1}$ (1.02-1.27)	*.02*	$1.12_{0,0}$ (0.68-1.83)	.65
Perceived usefulness of the app	$0.82_{2,1}$ (0.68-0.99)	*.04*	—	—
Alcohol consumption	—	—	—	—
Perceived lack of time	$0.93_{2,1}$ (0.86-1.02)	.15	—	—
P2				
Reminder	$1.99_{1,0}$ (0.67-5.94)	.22	—	—
Motivation to reduce alcohol	—	—	—	—
Perceived usefulness of the app	—	—	—	—
Alcohol consumption	$1.50_{1,0}$ (1.16-1.93)	*.003*	$2.38_{1,0}$ (1.65-3.43)	*<.001*
Perceived lack of time	$1.13_{1,0}$ (1.01-1.25)	*.03*	—	—
P3				
Reminder	—	—	$4.31_{0,0}$ (1.73-10.73)	*.003*
Motivation to reduce alcohol	$0.89_{1,0}$ (0.67-1.19)	.45	—	—
Perceived usefulness of the app	—	—	—	—
Alcohol consumption	—	—	$1.38_{0,0}$ (1.11-1.73)	*.006*
Perceived lack of time	—	—	$1.19_{0,0}$ (0.79-1.77)	.40
P4[e]				
Reminder	—	—	—	—
Motivation to reduce alcohol	$1.88_{0,0}$ (1.22-2.91)	*.005*	$2.03_{0,0}$ (1.72-2.40)	*<.001*
Perceived usefulness of the app	—	—	$137.33_{0,0}$ (49.45-381.34)	*<.001*
Alcohol consumption	—	—	—	—
Perceived lack of time	—	—	$0.20_{0,0}$ (0.14-0.29)	*<.001*
P5[f]				
Motivation to reduce alcohol	—	—	—	—
Perceived usefulness of the app	$1.42_{2,2}$ (1.15-1.75)	*.002*	$1.39_{0,0}$ (1.06-1.82)	*.02*
Alcohol consumption	—	—	—	—
Perceived lack of time	$1.08_{2,2}$ (0.81-1.43)	.60	—	—
P6				
Reminder	$3.88_{2,0}$ (1.37-11.03)	*.01*	—	—
Motivation to reduce alcohol	$1.07_{2,0}$ (0.93-1.21)	.35	$3.45_{0,0}$ (1.34-8.83)	*.01*
Perceived usefulness of the app	$1.12_{2,0}$ (0.94-1.34)	.21	—	—
Alcohol consumption	$0.92_{2,0}$ (0.83-1.02)	.13	—	—
Perceived lack of time	$0.77_{2,0}$ (0.61-0.97)	*.03*	$1.24_{0,0}$ (0.71-2.17)	.45
P7				
Reminder	$3.26_{1,0}$ (2.15-4.96)	*<.001*	—	—

Participant	Frequency of engagement[a]		Amount of engagement[a]	
	Incidence rate ratio (IRR) (95% CI)	*P* value	IRR (95% CI)	*P* value
Motivation to reduce alcohol	—	—	$1.67_{0,0}$ (1.16-2.40)	*.008*
Perceived usefulness of the app	—	—	$0.52_{0,0}$ (0.33-0.80)	*.005*
Alcohol consumption	—	—	—	—
Perceived lack of time	—	—	—	—
P8[f]				
Motivation to reduce alcohol	—	—	—	—
Perceived usefulness of the app	—	—	—	—
Alcohol consumption	$0.85_{1,0}$ (0.67-1.09)	.20	$0.82_{0,0}$ (0.47-1.43)	.50
Perceived lack of time	—	—	$1.33_{0,0}$ (0.97-1.82)	.08
P9				
Reminder	—	—	—	—
Motivation to reduce alcohol	—	—	$1.20_{1,1}$ (0.92-1.58)	0.18
Perceived usefulness of the app	$1.38_{1,0}$ (1.24-1.53)	*<.001*	$1.67_{1,1}$ (1.22-2.29)	*.002*
Alcohol consumption	—	—	—	—
Perceived lack of time	—	—	$4.77_{1,1}$ (1.09-20.79)	*.04*

[a]All models were adjusted for the day of the week using a cyclic cubic smoothing term.

[b]Numbers in subscript indicate the lags of autoregressive (AR) and moving average (MA) terms, respectively. A lag value of 0 indicates that an AR or an MA term was not included.

[c]*P* values significant at the .05 level are highlighted in italics.

[d]Indicates that a predictor variable was not included in the best-fitting model.

[e]For P4, generalized additive mixed models would not converge. Therefore, generalized additive models were fitted.

[f]As P5 and P8 opted out of receiving the daily reminder, this variable did not apply to these 2 participants.

Daily Reminder

In univariable analyses, the daily reminder was a significant predictor of the frequency of engagement for 3 participants (P1, P7, and P9; see Figure 1). In multivariable analyses, the daily reminder was a significant predictor for 3 participants (IRR=1.80-3.88; all *P*<.05). For these participants (P1, P6, and P7), the receipt of a reminder was associated with an 80% to 288% increase in the number of log-ins in the next 12 hours (see Table 6).

In univariable analyses, the daily reminder was a significant predictor of the amount of engagement for 3 participants (P3, P6, and P7; see Figure 2). In multivariable analyses, the daily reminder was a significant predictor for 1 participant (IRR=4.31; 95% CI 1.73-10.73; *P*<.01). For this participant (P3), the receipt of a reminder was associated with a 331% increase in the amount of engagement in the next 12 hours (see Table 6).

Figure 1. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of the daily reminder and the frequency of engagement for each participant (y-axis) in univariable analyses. The vertical line indicates parity; 95% CIs that cross the line of parity indicate nonsignificant incidence rate ratios. For P4, the univariable model did not converge. P4 is hence not included in this plot. P: participant.

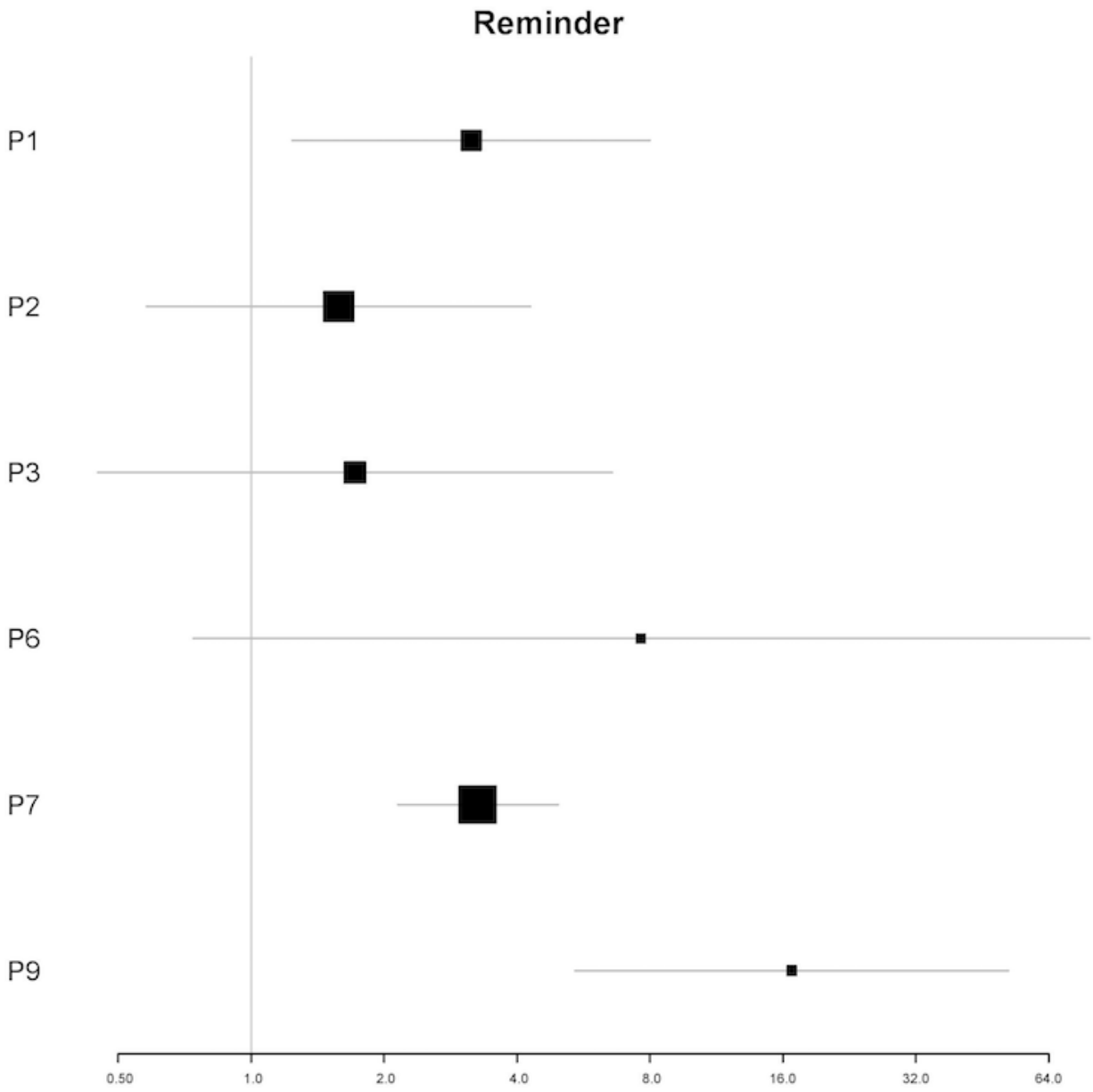
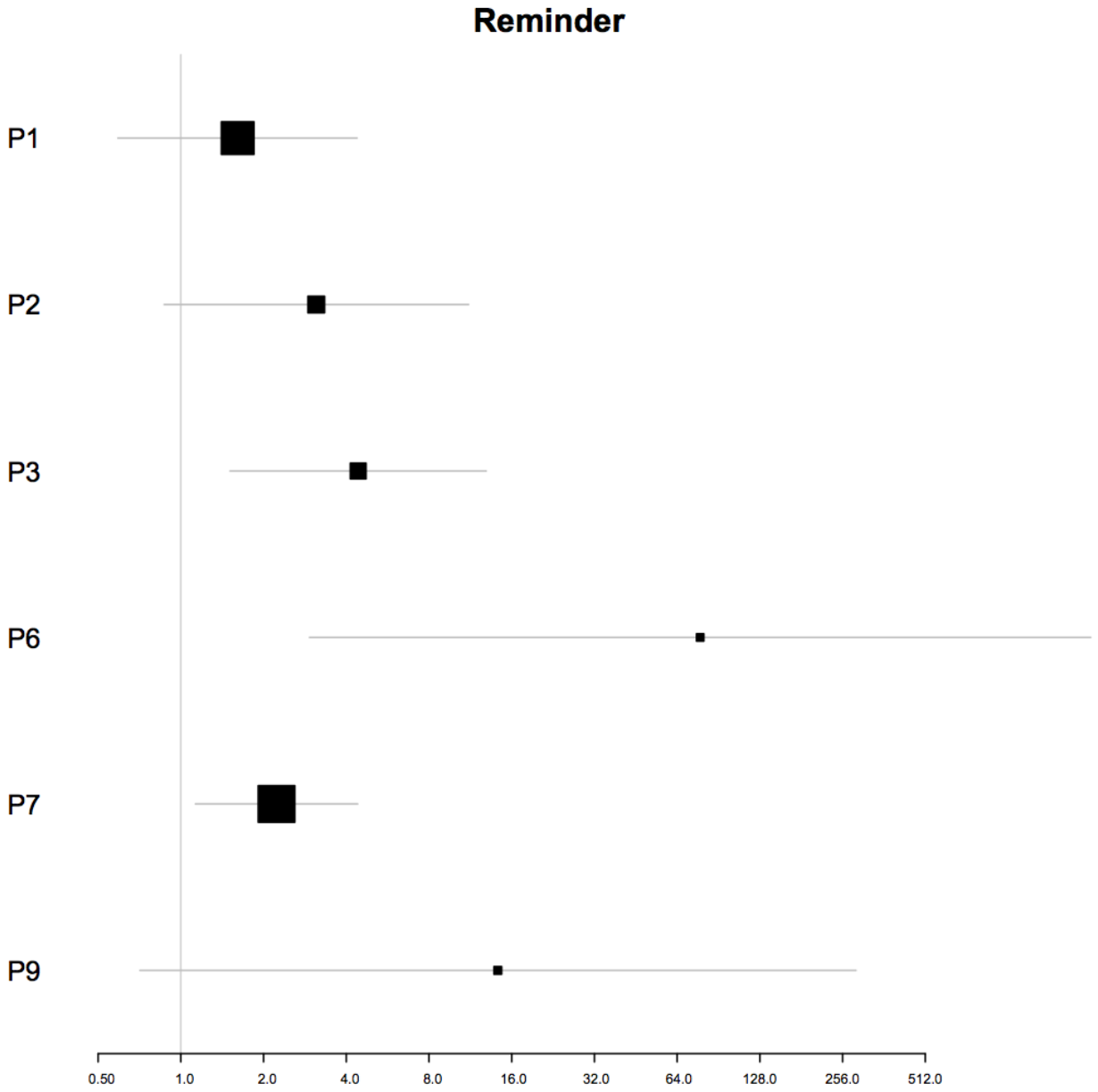


Figure 2. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of the daily reminder and the amount of engagement for each participant (y-axis) in univariable analyses. For P4, the univariable model did not converge. P4 is hence not included in this plot. P: participant.



Motivation to Reduce Alcohol

In univariable analyses, motivation to reduce alcohol was a significant predictor of the frequency of engagement for 2 participants (P4 and P6; see Figure 3). In multivariable analyses, motivation to reduce alcohol was a significant predictor for 1 participant (IRR=1.14; 95% CI 1.02-1.27; *P*=.02). For this participant (P4), a 1-point increase in motivation to reduce alcohol was associated with a 14% increase in the number of log-ins in the next 12 hours (see Table 6).

In univariable analyses, motivation to reduce alcohol was a significant predictor of the amount of engagement for 3 participants (P4, P6, and P9; see Figure 4). In multivariable analyses, motivation to reduce alcohol was a significant predictor for 3 participants (IRR=1.67-3.45; all *P*<.05). For these participants (P4, P6, and P7), a 1-point increase in motivation was associated with a 67% to 245% increase in the amount of engagement in the next 12 hours (see Table 6).

Figure 3. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of motivation to reduce alcohol and the frequency of engagement for each participant (y-axis) in univariable analyses. P: participant.

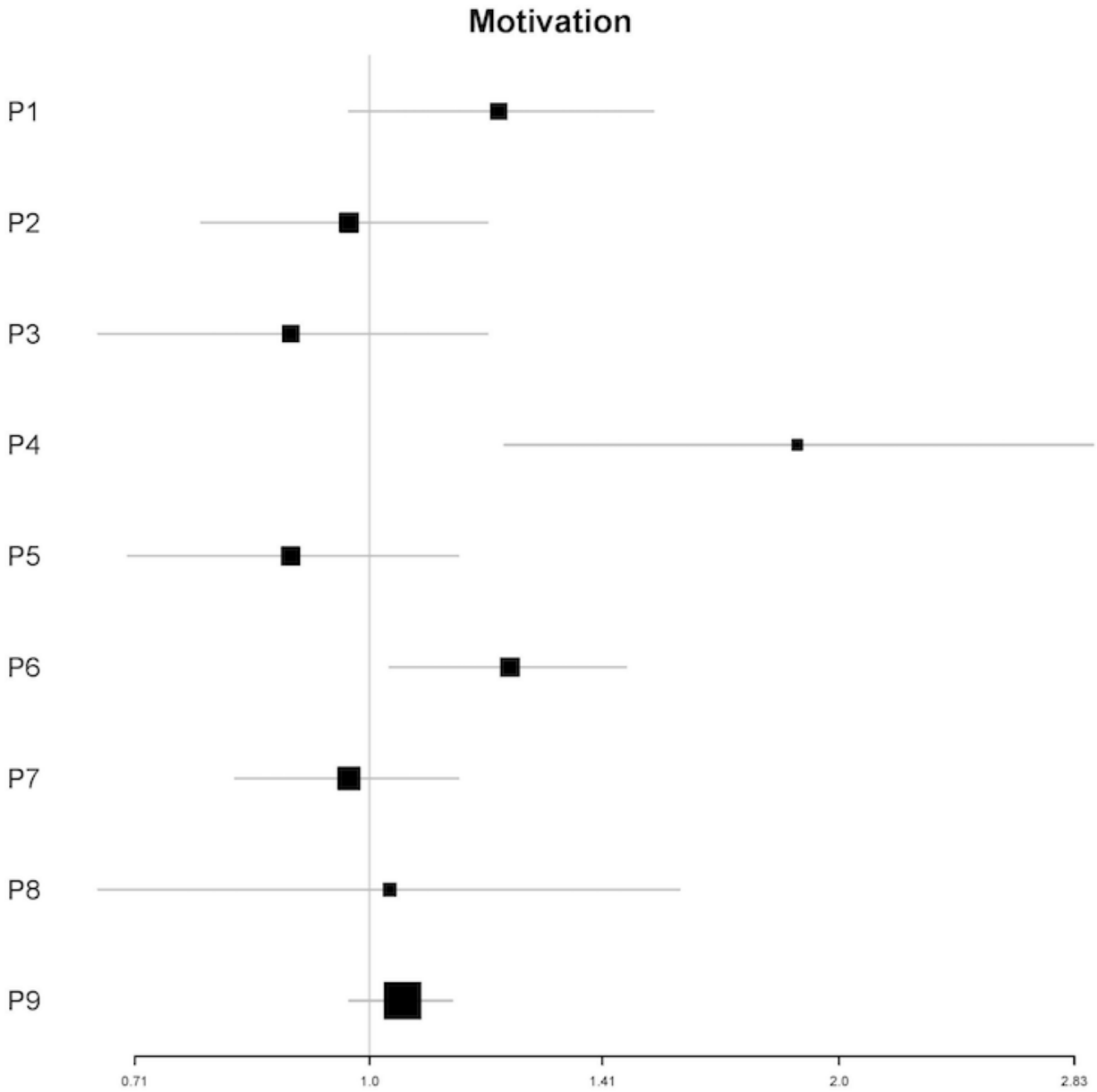
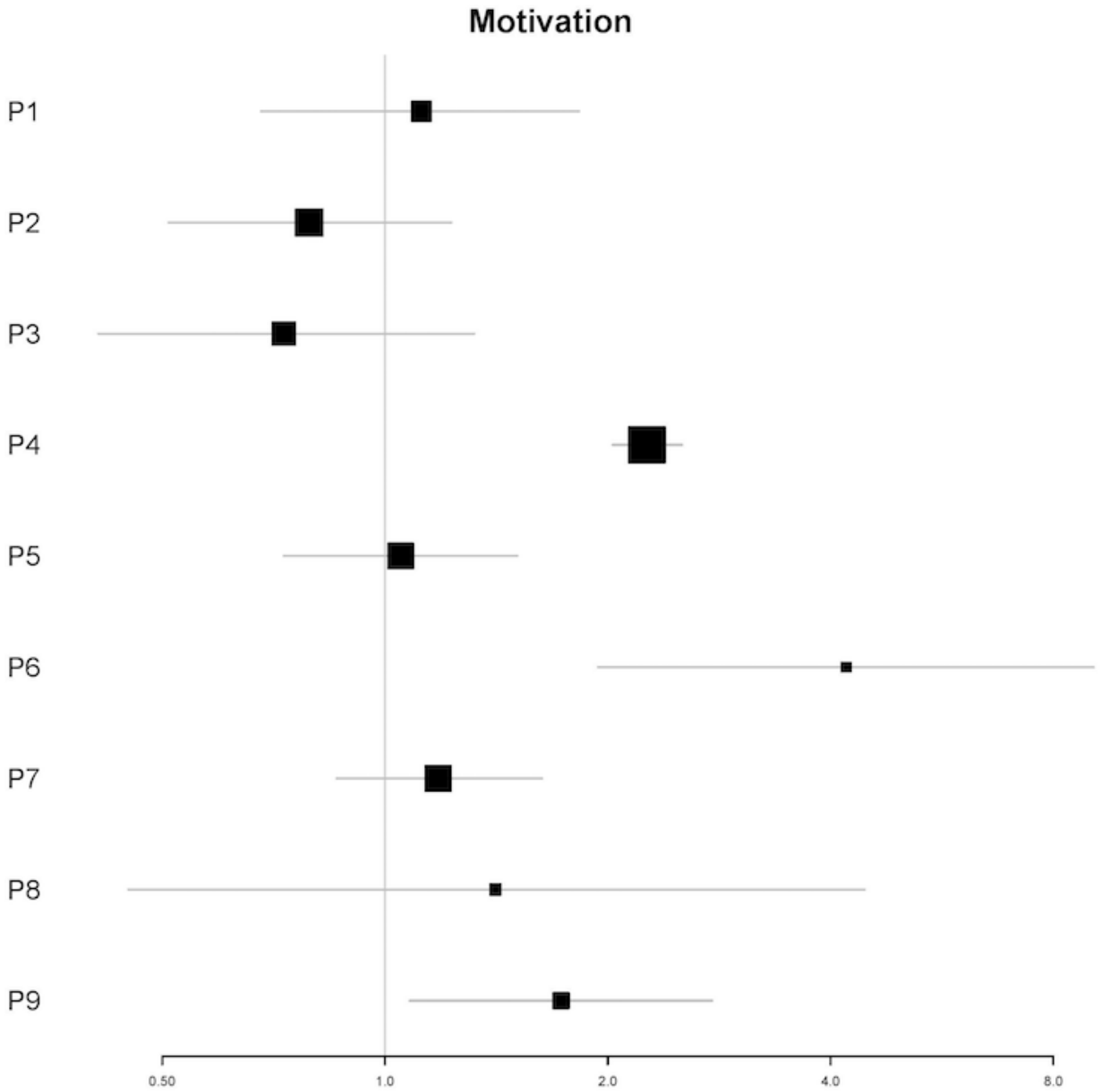


Figure 4. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of motivation to reduce alcohol and the amount of engagement for each participant (y-axis) in univariable analyses. P: participant.



Perceived Usefulness of the App

In univariable analyses, the perceived usefulness of the app was a significant predictor of the frequency of engagement for 3 participants (P4, P6, and P9; see Figure 5). In multivariable analyses, perceived usefulness of the app was a significant predictor for 3 participants (IRR=0.82-1.42; all $P<.05$). For 1 participant (P1), a 1-point increase in the perceived usefulness of the app was associated with an 18% reduction in the number of log-ins in the next 12 hours, whereas for 2 participants (P5 and P9), a 1-point increase in the perceived usefulness of the app was associated with a 38% to 42% increase in the number of log-ins in the next 12 hours (see Table 6).

In univariable analyses, the perceived usefulness of the app was a significant predictor of the amount of engagement for 3 participants (P4, P5, and P9; see Figure 6). In multivariable analyses, the perceived usefulness of the app was a significant predictor for 4 participants (IRR=0.52-137.32; all $P<.05$). For 1 participant (P7), a 1-point increase in the perceived usefulness of the app was associated with a 48% reduction in the amount of engagement in the next 12 hours. For 3 participants (P4, P5, and P9), a 1-point increase in perceived usefulness of the app was associated with a 67% to 13,632% increase in the amount of engagement in the next 12 hours (see Table 6).

Figure 5. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of perceived usefulness of the app and the frequency of engagement for each participant (y-axis) in univariable analyses. P: participant.

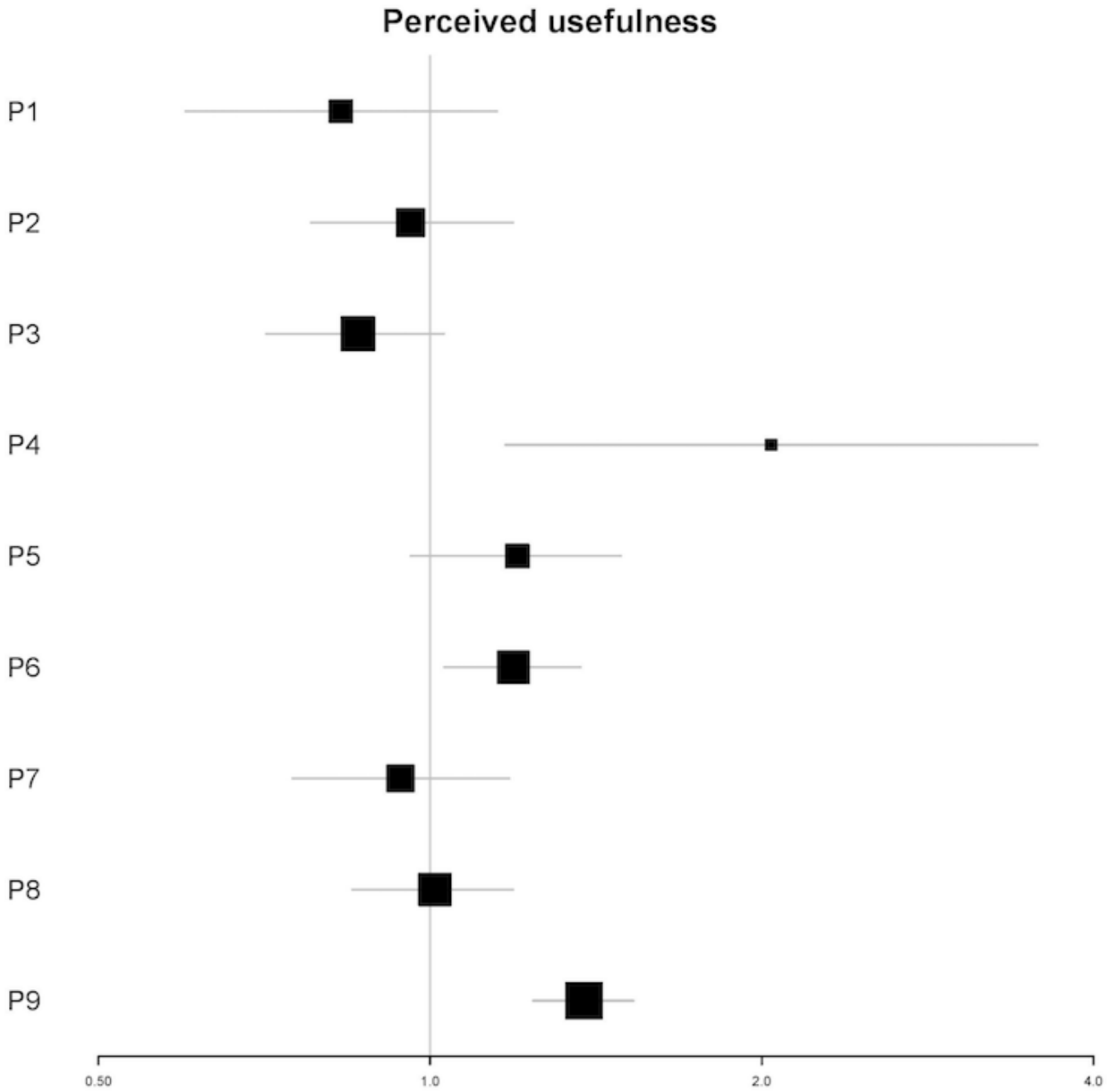
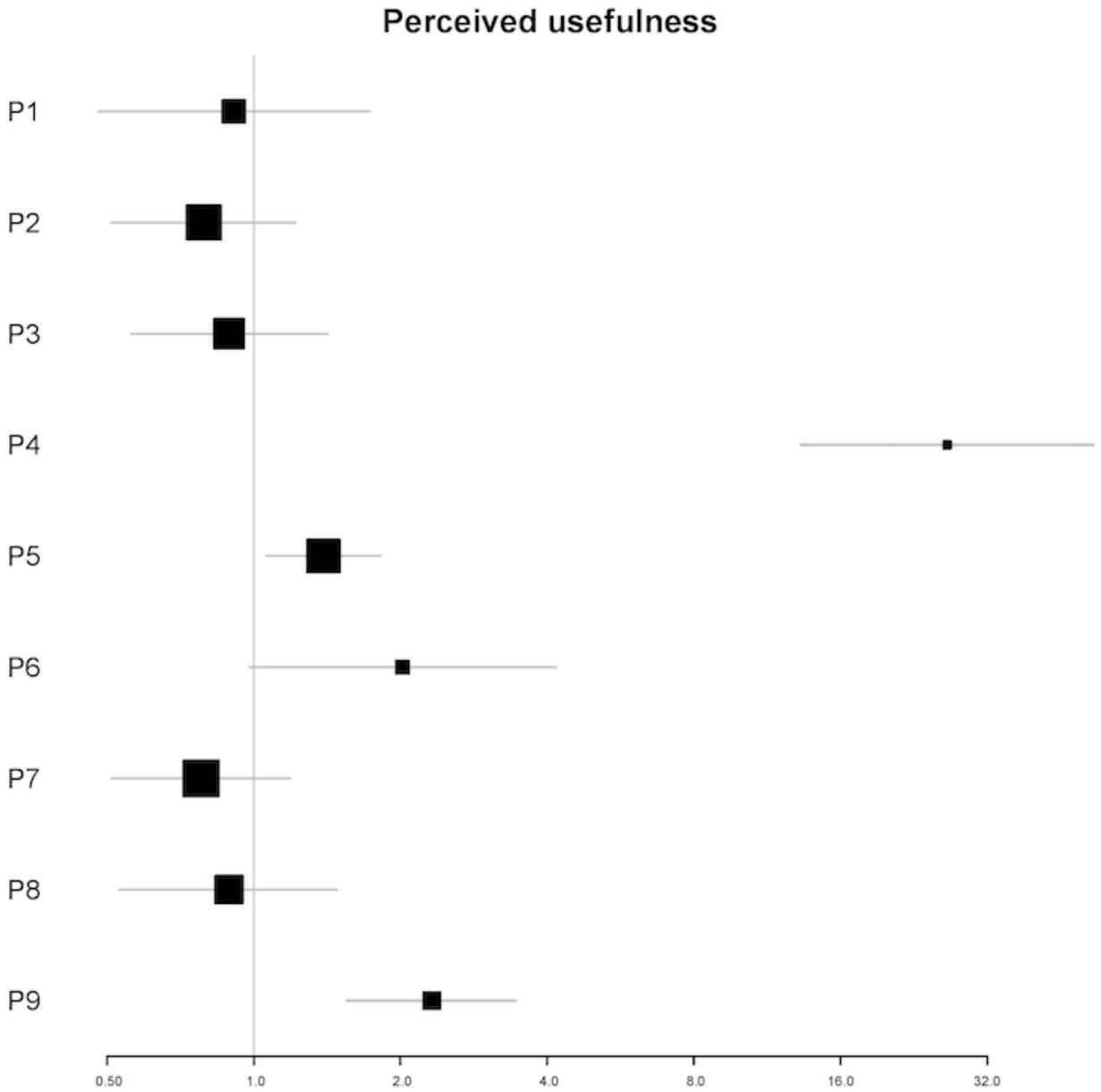


Figure 6. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of perceived usefulness of the app and the amount of engagement for each participant (y-axis) in univariable analyses. P: participant.



Alcohol Consumption

In univariable analyses, the number of drinks containing alcohol consumed in the past 12 hours was a significant predictor of the frequency of engagement for 1 participant (P2; see Figure 7). In multivariable analyses, the number of drinks containing alcohol consumed in the past 12 hours was a significant predictor for 1 participant (IRR=1.50; 95% CI 1.16-1.93; *P*<.01). For this participant (P2), each alcoholic drink consumed in the past 12 hours was associated with a 50% increase in the number of log-ins in the next 12 hours (see Table 6).

In univariable analyses, the number of drinks containing alcohol consumed in the past 12 hours was a significant predictor of the amount of engagement for 2 participants (P2 and P3; see Figure 8). In multivariable analyses, the number of drinks containing alcohol consumed in the past 12 hours was a significant predictor for 2 participants (IRR=1.38-2.38; *P*<.01). For these participants (P2 and P3), each alcoholic drink consumed in the past 12 hours was associated with a 38% to 138% increase in the amount of engagement in the next 12 hours (see Table 6).

Figure 7. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of alcohol consumption and the frequency of engagement for each participant (y-axis) in univariable analyses. For P4, the univariable model did not converge. P4 is hence not included in this plot. P: participant.

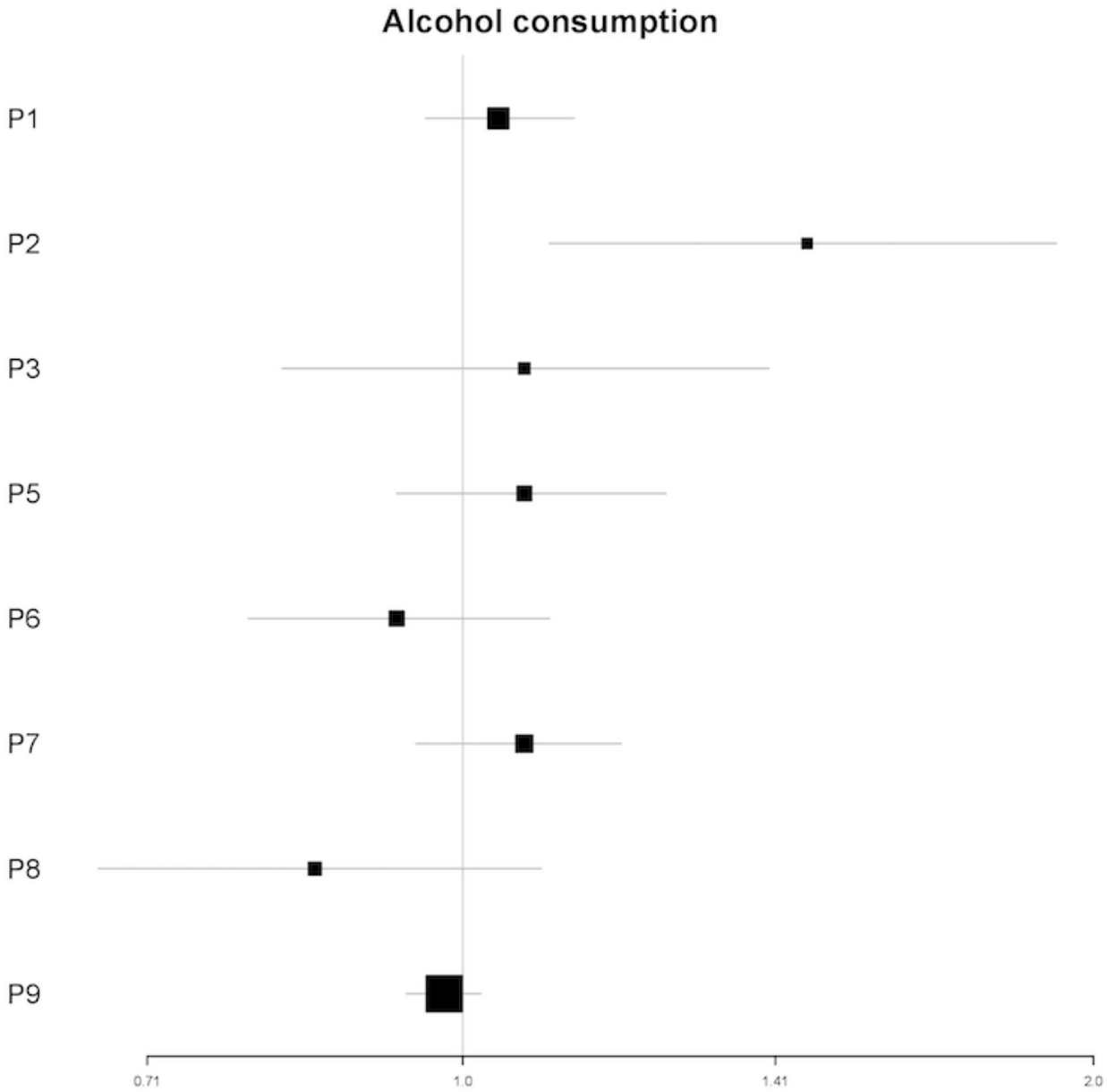
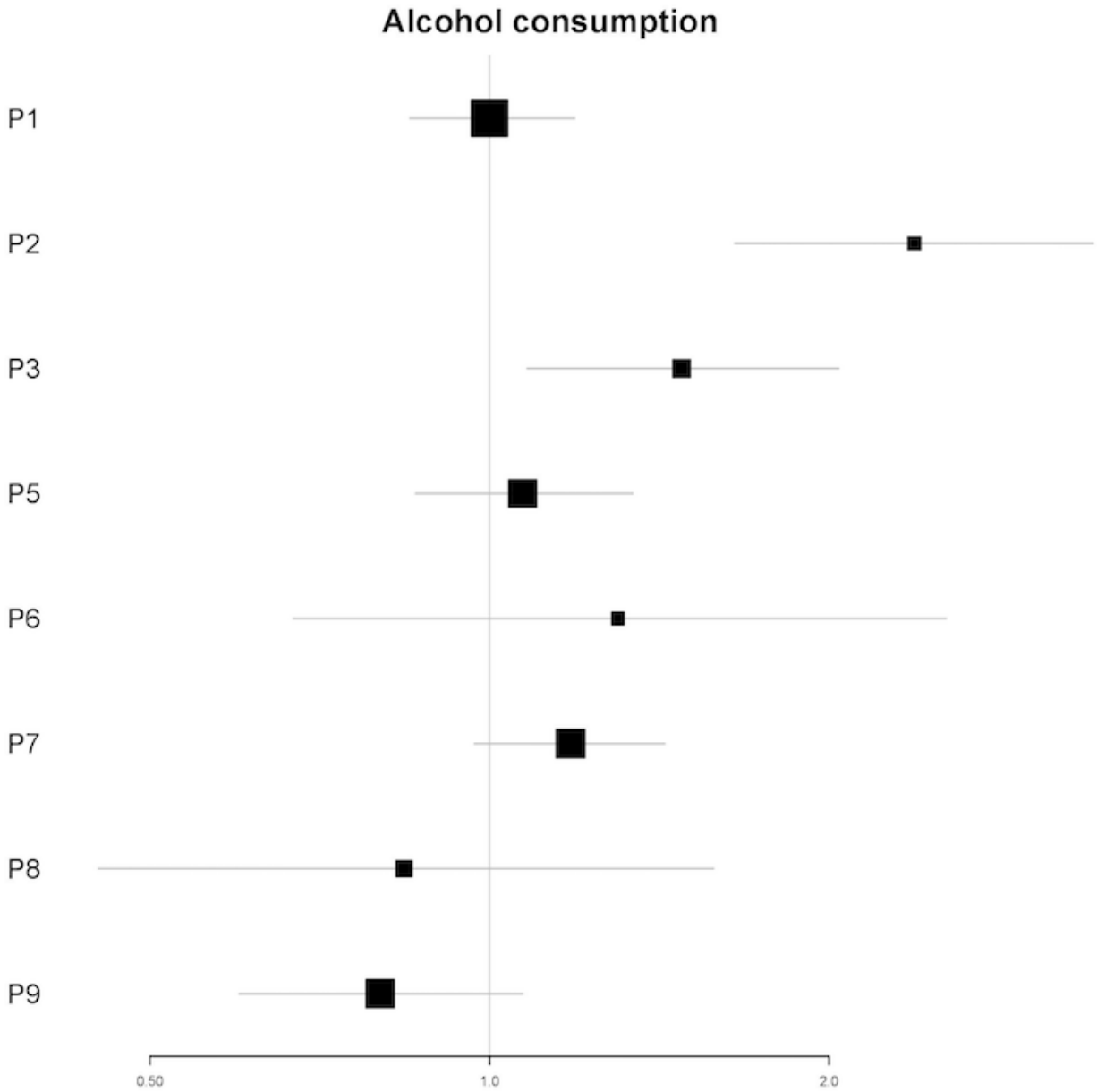


Figure 8. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of alcohol consumption and the amount of engagement for each participant (y-axis) in univariable analyses. For P4, the univariable model did not converge. P4 is hence not included in this plot. P: participant.



Perceived Lack of Time

In univariable analyses, the perceived lack of time was not a significant predictor of the frequency of engagement for any of the participants (see Figure 9). In multivariable analyses, the perceived lack of time was a significant predictor for 2 participants (IRRs=0.77-1.13; *P*<.05). For 1 participant (P6), a 1-point increase in the perceived lack of time (meaning that they had more time for the app) was associated with a 23% reduction in the number of log-ins in the next 12 hours. For the other participant (P2), a 1-point increase in the perceived lack of time was associated with a 13% increase in the number of log-ins in the next 12 hours (see Table 6).

In univariable analyses, the perceived lack of time was a significant predictor of the amount of engagement for 4 participants (P1, P4, P6, and P9; see Figure 10). In multivariable analyses, the perceived lack of time was a significant predictor for 2 participants (IRRs=0.20-4.77; *P*<.05). For 1 participant (P4), a 1-point increase in the perceived lack of time (meaning that they had more time for the app) was associated with an 80% reduction in the amount of engagement in the next 12 hours. For the other participant (P9), a 1-point increase in the perceived lack of time was associated with a 377% increase in the amount of engagement in the next 12 hours (see Table 6).

Figure 9. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of perceived lack of time and the frequency of engagement for each participant (y-axis) in univariable analyses. P: participant.

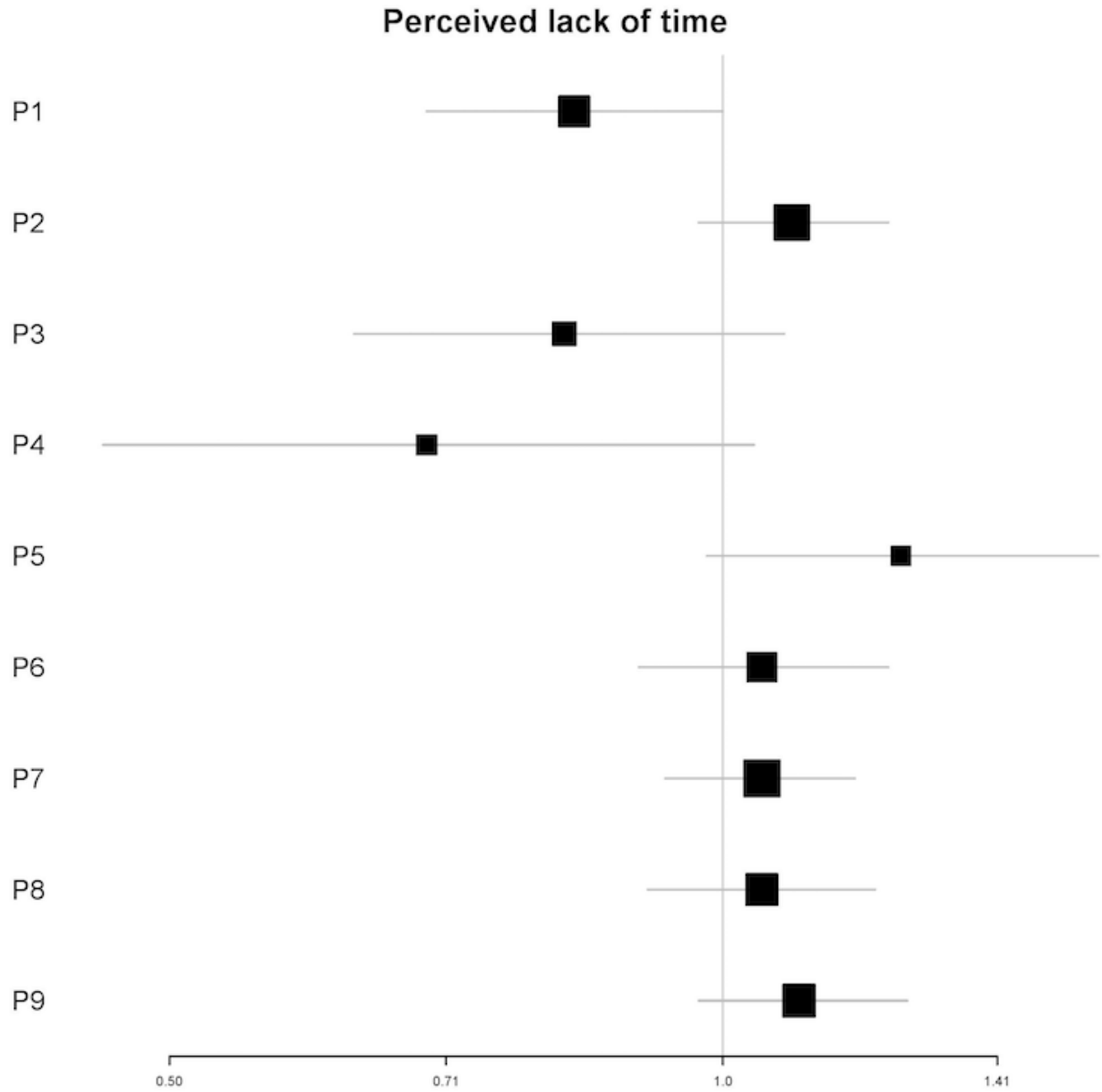
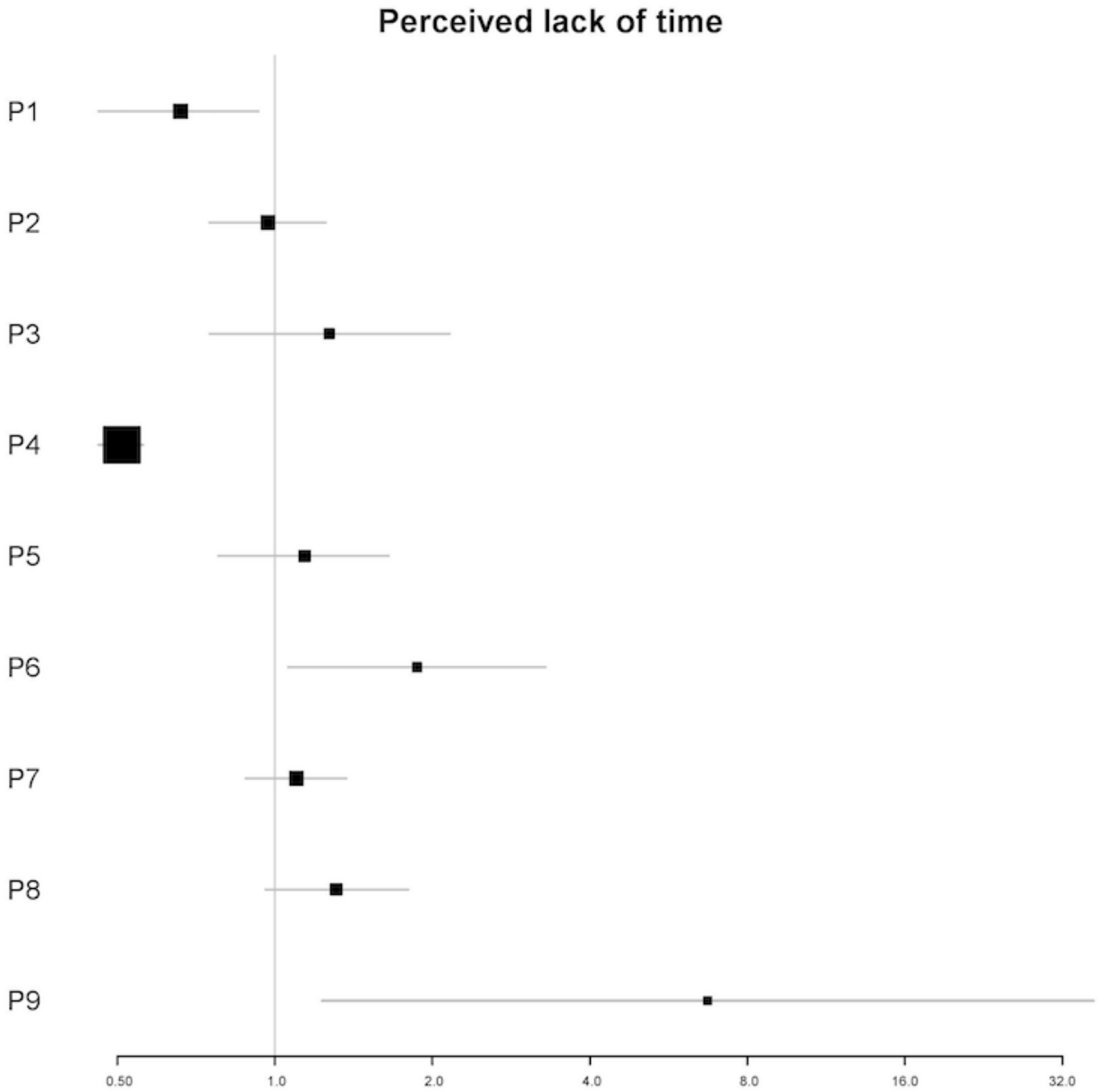


Figure 10. Plot of incidence rate ratios and 95% CIs (x-axis) for the association of perceived lack of time and the amount of engagement for each participant (y-axis) in univariable analyses. P: participant.



Debriefing Interviews

Establishing a Routine

When asked to reflect on their engagement with the *Drink Less* app, the majority of participants (P2, P3, P5, P6, P7, and P8) mentioned that they established a routine to engage with the app on a daily basis over the 28-day study. They would, for example, remember to open the app every morning upon waking or when traveling to work, or every evening when returning home after work. Some participants (who had opted to receive the daily push notification) thought this was facilitated by the daily reminder:

> *I've sort of made a habit of it now, and [I'm] probably going to continue as well.* [P2]

> *I was using it every day, because I just wanted to put the summary in for the day, even if it was a drink free day. So I would always use it.* [P8]

Purposeful Versus Purposeless Engagement

The majority of participants (P1, P3, P6, P7, P8, P9, and P10) reported that they quickly learned which features they "had to" engage with. They would only open the app for a specific purpose, which typically involved logging drinks or alcohol-free days in the calendar and reviewing their progress on the dashboard, as opposed to opening the app for entertainment:

> *I can just go on, quickly, input the stuff, have a check of how I'm doing against the target, and then go off it.* [P7]

Momentary Triggers and Barriers to Engagement

Most participants did not feel inclined to open the app when they were in a social setting, not necessarily because they anticipated feeling embarrassed if friends, family, or colleagues would ask about why they were using an alcohol reduction app but because they wanted to stay focused on their interactions with other people:

> *Not necessarily just because like: "Oh, I don't want them to know that I'm doing it", more just like, "I'm busy and I'm having a good time, and I'll do it later."* [P7]

Some participants (P4, P5, P7, and P9) mentioned that they thought they were more likely to open the app when feeling bored. A participant (P2) tended to open the app to combat momentary cravings to drink. Some participants (P2, P7, P9, and P10) thought they were less likely to use the app when they were hungover or experiencing low mood:

> *I'd sort of open the game to distract myself, and say that I should not be saying yes to everything.* [P2]

Discussion

Principal Findings

This series of *N*-of-1 studies found that the utility of app-related and psychological variables in predicting 2 facets of behavioral engagement (ie, the frequency and the amount of engagement) with an alcohol reduction app differed within and between individuals. This suggests that different strategies to promote engagement may be required for different individuals, and that such strategies may have differential effects on the various facets of engagement.

In line with findings from group-level studies [47], the receipt of a proactive reminder was significantly associated with the frequency of engagement for a few participants. However, this was not the case for all participants who had opted to have the reminder switched on. This suggests that some participants may be more responsive to prompts than others. However, for some participants, significant associations were only observed in the multivariable (and not in the univariable) analyses. As this may reflect suppression effects, results for participants with inconsistent associations across uni- and multivariable analyses should be interpreted with caution. For participants receiving the daily reminder in the middle of a 12-hour measurement period (eg, P3), it was not possible to assess whether the receipt of the reminder occurred before or after app engagement, as all predictor variables were entered into one multivariable model.

In contrast to results from group-level studies [13,15], motivation to reduce alcohol was significantly associated with the amount, but not necessarily the frequency, of engagement for some participants. For these individuals, being more highly motivated to reduce alcohol consumption may make them more willing to spend time (and perhaps also effort) on the app, provided that they have decided to open the app in the first place.

Previous group-level studies have identified a negative relationship of baseline alcohol consumption with the frequency of engagement, such that the higher the alcohol consumption, the less frequent the engagement [12,13,16]. In this study, none of the participants engaged with the app at a lower rate after sessions of heavier alcohol consumption. Instead, alcohol consumption was positively related to the frequency and the amount of engagement for some participants. It is plausible that the direction of the relationship between engagement and the target behavior may vary across individuals: while some participants may be more prone to engage when they are doing well (ie, having abstained from or consumed less-than-typical amounts of alcohol), the reverse may hold for other participants.

The variable 'perceived lack of time' has typically been explored qualitatively in interviews with participants who have dropped out of RCTs of digital interventions [13]. For some participants in the present study, this variable was significantly associated with the frequency and the amount of engagement. However, the direction of the relationships varied across participants, with some participants displaying lower rates of engagement after having indicated that they had a lot of time available for the app. It should, however, be noted that for some participants (ie, P2 and P6), significant associations were only observed in the multivariable analyses. Hence, results for these participants should be interpreted with caution. P4 (who displayed significant negative associations across both uni- and multivariable analyses) may have rated herself as having a lot of time for the app at the time of the morning or evening survey, but this might have changed a few hours later, which might have interfered with app use. More frequent EMAs may, therefore, help to detect a relationship between perceived lack of time and engagement for some participants. Alternatively, participants' availability/receptivity to engage could be automatically inferred from their calendar or phone activity [48].

In line with findings from group-level studies [41,49], the variable 'perceived usefulness of the app' was found to be one of the most consistent predictors of both the frequency and the amount of engagement with the *Drink Less* app. The direction of the associations differed across participants; although the perceived usefulness of the app tended to be positively associated with the frequency and the amount of engagement, the reverse was observed for some participants. Again, this might be indicative of the need to capture this variable at a higher resolution (ie, more frequent EMAs). Alternatively, this variable may have been subject to social desirability. It should also be noted that for some participants (ie, P1 and P7), significant associations were only observed in the multivariable analyses.

For some participants, none of the variables assessed were significantly associated with the frequency (ie, P3 and P8) or the amount of engagement (ie, P8) in either the uni- or multivariable analyses. This raises the question as to what was driving engagement for these participants. A plausible explanation in relation to frequency, as mentioned in the debriefing interviews, is that these participants established a routine to engage with the app. If this was indeed the case, habit formation could be tested as a promising strategy to promote engagement for other users [50]. The debriefing interviews were unable to shed light on key factors that might have driven participants' amount of engagement because it was difficult for

them to introspect about momentary influences on time spent on the app (particularly as the time unit of interest was seconds rather than minutes or hours). It should be noted that although daily engagement with alcohol reduction apps, such as *Drink Less*, may be brief on average, thus making it difficult for users to introspect about momentary influences on their app use, this may not generalize to apps for other behaviors or activities. For example, apps for physical activity or mindfulness meditation, which have typically been designed to be kept open while performing the target behavior, may generate larger amounts of engagement. Hence, it may be easier for users to reflect on their daily engagement with such apps [51,52].

Strengths

To the authors' knowledge, this was the first study to examine within-person predictors of the frequency and the amount of engagement with an alcohol reduction app. The predictors assessed in this study were selected based on evidence from group-level studies and in-depth qualitative studies with potential users of alcohol reduction apps. Compliance with the twice-daily EMAs was high (0%-16% missing data), and the automatic recording of the outcome variables in real time ensured that participant burden and missing outcome data were minimized. This study provides initial evidence that it is feasible and acceptable to gather data in this manner and a novel time series approach (ie, GAMMs) can be successfully used to model data from *N*-of-1 studies.

Limitations

This study was conceptualized as a series of observational *N*-of-1 studies; however, participants engaged with an active digital intervention and study materials, which included behavior change techniques known to alter cognition and behavior (eg, prompts and self-monitoring) [53]. It is, therefore, possible that both predictor and outcome variables were subject to nonrandom fluctuations that were caused by participants' engagement with the intervention and study materials. However, as engagement with digital interventions cannot be studied in isolation, without asking participants to engage with a particular intervention and related study materials, it was not possible to overcome this particular limitation.

The study sample was almost exclusively women. As men tend to exhibit more alcohol-related problems than women [54,55], it is unclear whether the same patterns of results would be observed in a more balanced or male-dominated sample. None of the participants dropped out of the study, suggesting that they were highly motivated to take part in the research. It is, therefore, possible that different patterns of results may be obtained in samples of less committed participants. The *Drink Less* app is currently available for iOS only. As market research suggests that iPhone users tend to be more affluent than Android users [56], different patterns of engagement might be observed in a sample of Android users. Participants were all aged younger than 32 years; older adults may display different patterns of engagement. It should, however, be noted that the aim of this study was not to produce results that are generalizable at the group level. In addition, 1 participant (P7) had used the app before the study period, which may have influenced their engagement. However, as participants serve as their own controls in *N*-of-1 studies, the finding that P7 engaged more frequently with the app when she had received the daily reminder is a meaningful piece of information; it could be used to inform the development of personalized engagement strategies for this unique user.

To keep participant burden to a minimum, other facets of engagement during each log-in session (eg, attention and interest) were not assessed. This study was, therefore, unable to highlight potentially interesting relationships between the predictor variables and experiential engagement ([6]; Perski et al, in press). Moreover, many participants opted to be reminded during the first measurement period (ie, during the day). As there is more time for engagement in the daytime (as compared with the nighttime), this may have confounded the observed relationship between the receipt of the daily reminder and the frequency and amount of engagement.

Avenues for Future Research

Descriptive plots were used to summarize the associations of each predictor variable with the key outcome variables across participants; it was not possible to pool results from the multivariable models in a meta-analysis. As time series analysis is becoming increasingly popular in the context of *N*-of-1 studies, suitable meta-analytic techniques are evolving [44], and this should be considered in future research. For studies with a greater number of participants, multilevel models (including the GAMM) can be used to estimate both within- and between-person effects [57].

Future research should test the feasibility of using both time- and event-prompted EMAs, with participants being prompted to respond to a few questions about their experiential engagement immediately after having opened the app. This would require careful piloting given the additional participant burden and unpredictability of response requests: it is possible that this might create a disincentive to open the app as participants may anticipate an additional cost directly linked to doing so. As indicated in the debriefing interviews, it is plausible that participants' physical location (eg, being in a social setting) is negatively associated with behavioral engagement for some participants. This could be explored further by means of accessing the location sensing data from participants' smartphones.

The feasibility and utility of just-in-time adaptive interventions (JITAIs) [58] for promoting engagement with alcohol reduction apps should be explored further. The JITAI is a type of intervention that is specifically designed to address the dynamically changing needs of individuals. JITAIs use inputs from, for example, EMAs or data collected via wearables or the phone's location sensors to inform what type of support each individual might need in different situations or contexts. They then automatically trigger support when the system infers that the individual is in need of or most receptive to that support. In the context of the results from this study, a JITAI could, for example, be delivered when an individual's level of perceived usefulness of the app or motivation to reduce alcohol is inferred to be below a given threshold for action, with a view to promoting the frequency of engagement.

Future research should consider the use of observational or experimental *N*-of-1 study designs as a valuable part of intervention development. Results from this study are currently being used to inform the optimization of the *Drink Less* app, involving, for example, the optimization of the content and timing of the daily reminder, with a view to promoting engagement.

Conclusions

This series of *N*-of-1 studies found that the utility of psychological and app-related variables in predicting the frequency and the amount of engagement with an alcohol reduction app differed within and between individuals. This highlights that important within-person associations may be masked in group-level designs and suggests that different strategies to promote engagement may be required for different individuals.

Acknowledgments

The authors would like to thank the participants for their time. OP is funded by Bupa under its partnership with UCL and Cancer Research UK (C1417/A22962). CG and EB receive salary support from Cancer Research UK (C1417/A22962). The authors gratefully acknowledge all funding. They also acknowledge the members of UCL's Tobacco and Alcohol Research Group for providing feedback on an early draft of the paper.

Authors' Contributions

OP, FN, CG, AB, SM, EB, and RW designed the study. OP collected the data, conducted the statistical analyses, and wrote the first draft of the paper. All authors have contributed to the final version of the paper and agree with its submission to the *Journal of Medical Internet Research*.

Conflicts of Interest

EB has received unrestricted research funding from Pfizer. RW undertakes research and consultancy and receives fees for speaking from companies that develop and manufacture smoking cessation medications. OP, FN, CG, AB, and SM have no conflicts of interest to declare.

Multimedia Appendix 1
Plots of participants' frequency of engagement over the course of the study. The y-axis displays frequency counts (ie, the frequency of engagement per 12-hour measurement period); the x-axis displays the 56 measurement periods.
[PDF File (Adobe PDF File)551 KB - mhealth_v7i9e14098_app1.pdf]

Multimedia Appendix 2
Results from the univariable generalized additive mixed models.
[PDF File (Adobe PDF File)331 KB - mhealth_v7i9e14098_app2.pdf]

References

1. Gowing L, Ali R, Allsop S, Marsden J, Turf E, West R, et al. Global statistics on addictive behaviours: 2014 status report. Addiction 2015 Jun;110(6):904-919. [doi: 10.1111/add.12899] [Medline: 25963869]
2. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
3. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
4. Ritterband LM, Tate DF. The science of internet interventions. Introduction. Ann Behav Med 2009 Aug;38(1):1-3. [doi: 10.1007/s12160-009-9132-5] [Medline: 19816750]
5. Schueller S, Muñoz R, Mohr D. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec 3;22(6):478-483. [doi: 10.1177/0963721413495872]
6. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
7. Donkin L, Christensen H, Naismith S, Neal B, Hickie I, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

8. Yardley L, Spring B, Riper H, Morrison L, Crane D, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
9. 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]
10. Johnston D, Johnston M. Useful theories should apply to individuals. Br J Health Psychol 2013;18:469-473. [Medline: 23724956]
11. Molenaar P. A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. Measurement 2004 Oct;2(4):201-218. [doi: 10.1207/s15366359mea0204_1]
12. Linke S, Murray E, Butler C, Wallace P. Internet-based interactive health intervention for the promotion of sensible drinking: patterns of use and potential impact on members of the general public. J Med Internet Res 2007;9(2):e10 [FREE Full text] [doi: 10.2196/jmir.9.2.e10] [Medline: 17513281]
13. Postel M, de Haan HA, ter Huurne ED, van der Palen J, Becker E, de Jong CA. Attrition in web-based treatment for problem drinkers. J Med Internet Res 2011 Dec 27;13(4):e117 [FREE Full text] [doi: 10.2196/jmir.1811] [Medline: 22201703]
14. Garnett C, Perski O, Tombor I, West R, Michie S, Brown J. Predictors of engagement, response to follow-up and extent of alcohol reduction in users of a smartphone app, Drink Less: secondary analysis of a factorial randomized controlled trial. JMIR Mhealth Uhealth 2018;6(12):e11175 [FREE Full text] [doi: 10.2196/11175] [Medline: 30552081]
15. Chiauzzi E, Green TC, Lord S, Thum C, Goldstein M. My student body: a high-risk drinking prevention web site for college students. J Am Coll Health 2005;53(6):263-274 [FREE Full text] [doi: 10.3200/JACH.53.6.263-274] [Medline: 15900990]
16. Radtke T, Ostergaard M, Cooke R, Scholz U. Web-based alcohol intervention: study of systematic attrition of heavy drinkers. J Med Internet Res 2017;19(6):e217 [FREE Full text] [doi: 10.2196/jmir.6780] [Medline: 28659251]
17. Kelders S, Kok R, Ossebaard H, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
18. Perski O, Blandford A, Ubhi H, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
19. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
20. Perski O, Baretta D, Blandford A, West R, Michie S. Engagement features judged by excessive drinkers as most important to include in smartphone applications for alcohol reduction: a mixed-methods study. Digit Health 2018;4:1-15. [doi: 10.1177/2055207618785841]
21. Sedgwick P. What is an. Br Med J 2014;348:g2674. [doi: 10.1136/bmj.g2674] [Medline: 2014260588]
22. Nisbett R, Wilson T. Telling more than we can know: verbal reports on mental processes. Psychol Rev 1977 May;84(3):231-259. [doi: 10.1037/0033-295X.84.3.231]
23. Smyth J, Smyth JM. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003;4(1):35-52 [FREE Full text] [doi: 10.1023/A:1023657221954]
24. Stone A, Shiffman S. Ecological momentary assessment (EMA) in behavorial medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
25. McDonald S, Quinn F, Vieira R, O'Brien N, White M, Johnston DW, et al. The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. Health Psychol Rev 2017 Apr 25:1-17. [doi: 10.1080/17437199.2017.1316672] [Medline: 28406349]
26. Vieira R, McDonald S, Araújo-Soares V, Sniehotta F, Henderson R. Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies. Health Psychol Rev 2017;11(3):222-234. [doi: 10.1080/17437199.2017.1343680] [Medline: 28629262]
27. Garnett C, Crane D, West R, Brown J, Michie S. The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers. Transl Behav Med 2019;9(2):296-307 [FREE Full text] [doi: 10.1093/tbm/iby043] [Medline: 29733406]
28. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
29. Perski O, Naughton F, Garnett C, Blandford A, Beard E, West R, et al. Open Science Framework. 2018. Do daily fluctuations in psychological and app-related variables predict within-person variation in engagement with an alcohol reduction app? An N-of-1 study URL: https://osf.io/zn79m/
30. Babor T, Higgins-Biddle J, Saunders J, Monteiro M. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care, Second Edition. Geneva: World Health Organization; 2001.
31. Inauen J, Shrout PE, Bolger N, Stadler G, Scholz U. Mind the gap? An intensive longitudinal study of between-person and within-person intention-behavior relations. Ann Behav Med 2016 Dec;50(4):516-522. [doi: 10.1007/s12160-016-9776-x] [Medline: 26801785]

32. Duan N, Kravitz R, Schmid C. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. J Clin Epidemiol 2013;66:S21-S28 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.04.006] [Medline: 23849149]
33. Kwasnicka D, Dombrowski S, White M, Sniehotta F. N-of-1 study of weight loss maintenance assessing predictors of physical activity, adherence to weight loss plan and weight change. Psychol Health 2017;32(6):686-708. [doi: 10.1080/08870446.2017.1293057] [Medline: 28323457]
34. Kaner E, Beyer F, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017;CD011479(9) [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
35. Gilks W, Richardson S, Spiegelhalter D. Markov Chain Monte Carlo in Practice. Boca Raton, FL: CRC Press; 1995.
36. Quinn F, Johnston M, Johnston D. Testing an integrated behavioural and biomedical model of disability in N-of-1 studies with chronic pain. Psychol Health 2013;28(12):1391-1406. [doi: 10.1080/08870446.2013.814773] [Medline: 23863041]
37. Hobbs N, Dixon D, Johnston M, Howie K. Can the theory of planned behaviour predict the physical activity behaviour of individuals? Psychol Health 2013;28(3):234-249 [FREE Full text] [doi: 10.1080/08870446.2012.716838] [Medline: 22943555]
38. 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;3(2):e73 [FREE Full text] [doi: 10.2196/mhealth.3895] [Medline: 26123578]
39. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
40. Crane D, Garnett C, Brown J, West R, Michie S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies. Front Public Health 2017;5:39 [FREE Full text] [doi: 10.3389/fpubh.2017.00039] [Medline: 28421175]
41. King W, He J. A meta-analysis of the technology acceptance model. Inf Manage 2006 Sep;43(6):740-755 [FREE Full text] [doi: 10.1016/j.im.2006.05.003]
42. Google Analytics. 2017. How a web session is defined in Analytics URL: https://support.google.com/analytics/answer/2731565 [accessed 2018-02-06]
43. Qualtrics Survey Software. Provo, Utah; 2016. URL: http://www.qualtrics.com/ [accessed 2018-05-19]
44. Shadish WR, Zuur AF, Sullivan KJ. Using generalized additive (mixed) models to analyze single case designs. J Sch Psychol 2014;52(2):149-178. [doi: 10.1016/j.jsp.2013.11.004] [Medline: 24606973]
45. 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]
46. Braun A, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Vervloet M, Linn A, van Weert JC, de Bakker DH, Bouvy M, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
48. Hekler E, Rivera D, Martin C, Phatak S, Freigoun M, Korinek E. Tutorial for using control systems engineering to optimize adaptive mobile health interventions. J Med Internet Res 2018;20(6):e214. [Medline: 29954725]
49. Schepers J, Wetzels M. A meta-analysis of the technology acceptance model: investigating subjective norm and moderation effects. Inf Manage 2007 Jan;44(1):90-103 [FREE Full text] [doi: 10.1016/j.im.2006.10.007] [Medline: 25246403]
50. Beeken R, Leurent B, Vickerstaff V, Wilson R, Croker H, Morris S, et al. A brief intervention for weight control based on habit-formation theory delivered through primary care: results from a randomised controlled trial. Int J Obes 2017;41(2):246-254 [FREE Full text] [doi: 10.1038/ijo.2016.206] [Medline: 27867204]
51. Laurie J, Blandford A. Making time for mindfulness. Int J Med Inform 2016;96:38-50. [doi: 10.1016/j.ijmedinf.2016.02.010] [Medline: 26965526]
52. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
53. Wilding S, Conner M, Sandberg T, Prestwich A, Lawton R, Wood C, et al. The question-behaviour effect: a theoretical and methodological review and meta-analysis. Eur Rev Soc Psychol 2016;27(1):196-230. [doi: 10.1080/10463283.2016.1245940]
54. Wilsnack R, Vogeltanz N, Wilsnack S, Harris T, Ahlström S, Bondy S, et al. Gender differences in alcohol consumption and adverse drinking consequences: cross-cultural patterns. Addiction 2000;95(2):251-265. [Medline: 10723854]
55. Nolen-Hoeksema S, Hilt L. Possible contributors to the gender differences in alcohol use and problems. J Gen Psychol 2006;133(4):357-374. [doi: 10.3200/GENP.133.4.357-374] [Medline: 17128956]
56. Berg M. Statista - The Statistics Portal. 2014. iPhone Users Earn More URL: https://www.statista.com/chart/2638/top-line-platform-stats-for-app-usage-in-the-us/ [accessed 2019-01-28]

57. Kwasnicka D, Inauen J, Nieuwenboom W, Nurmi J, Schneider A, Short C, et al. Challenges and solutions for N-of-1 design studies in health psychology. Health Psychol Rev 2019;13(2):163-178 [FREE Full text] [doi: 10.1080/17437199.2018.1564627] [Medline: 30626274]
58. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014;3:17. [doi: 10.1109/MPRV.2014.46]

Abbreviations

AR: autoregressive
EMA: ecological momentary assessment
GAMM: generalized additive mixed model
IRR: incidence rate ratio
JITAI: just-in-time adaptive intervention
MA: moving average
RCT: randomized controlled trial
UCL: University College London

Edited by G Eysenbach; submitted 22.03.19; peer-reviewed by D Johnston, S Miller; comments to author 16.05.19; revised version received 23.05.19; accepted 11.06.19; published 02.10.19.

Please cite as:
Perski O, Naughton F, Garnett C, Blandford A, Beard E, West R, Michie S
Do Daily Fluctuations in Psychological and App-Related Variables Predict Engagement With an Alcohol Reduction App? A Series of N-Of-1 Studies
JMIR Mhealth Uhealth 2019;7(10):e14098
URL: https://mhealth.jmir.org/2019/10/e14098
doi:10.2196/14098
PMID:31579022

Original Paper

Young People's Attitudes and Motivations Toward Social Media and Mobile Apps for Weight Control: Mixed Methods Study

Charoula Konstantia Nikolaou[1], BSc, MSc, PhD; Zoey Tay[2], BSc, MSc; Jodie Leu[3], BSc, MSc; Salome Antonette Rebello[2], BSc, MSc, PhD; Lisa Te Morenga[4], BSc, MSc, PhD; Rob M Van Dam[2], BSc, MSc, PhD; Michael Ernest John Lean[5], BS, MD

[1]Graduate School of Public Health, Department of Biostatistics and Bioinformatics, St Luke's International University, Tokyo, Japan

[2]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[3]Research School of Population Health, College of Health and Medicine, Australian National University, Canberra, Australia

[4]Department of Human Nutrition, University of Otago, Dunedin, New Zealand

[5]Department of Human Nutrition, School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Charoula Konstantia Nikolaou, BSc, MSc, PhD
Graduate School of Public Health
Department of Biostatistics and Bioinformatics
St Luke's International University
3-6-2 Tsukiji, Chuo-ku
Tokyo, 104-0045
Japan
Phone: 81 08078059495
Email: cknikolaou1@gmail.com

Abstract

Background: Effective prevention at a young enough age is critical to halt the obesity epidemic. Mobile health (mHealth) apps would potentially reach large numbers at low-cost. While there is already a profusion of lifestyle apps, they are mostly non-evidence-based and evidently ineffective against rising obesity prevalence.

Objective: The aim of this study was to explore preferences and usage of lifestyle apps among young people in 6 countries.

Methods: A mixed methods study was conducted among young people aged 13 to 24 years residing in the United Kingdom, Belgium, Finland, Greece, Singapore, and New Zealand. Participants were recruited from Web advertisements on Facebook, asking for volunteers interested in mobile apps in general, not specific to lifestyle or health, to complete a short survey comprising 18 questions on demographics, weight gain, and mobile app preferences and then to join English-language online focus groups, which were held during 2017, in password-protected Web rooms, moderated by an experienced researcher. Descriptive statistics were carried out for the survey, and thematic analysis was applied to transcripts.

Results: A total of 2285 young people (610 adolescents aged 13-17 years and 1675 young adults aged 18-24 years) responded and completed the survey, with 72.0% (1645) reported being concerned about weight gain for themselves or friends. Later, 807 young people (376 adolescents and 431 young adults) were selected based on age and country to participate in 12 online focus groups, with 719 young people completing. Analysis revealed 4 main themes: (1) feelings toward personal weight; (2) perception of lifestyle apps and desired content for weight gain prevention; (3) social media apps, lifestyle apps, and motivation for downloading and retaining; and (4) data safety and data usage and confidentiality. Young people are interested in evidence-based advice in programs incorporating their preferences.

Conclusions: Young people are commonly, and consistently across 6 countries, concerned about weight gain and obesity and would welcome evidence-based mHealth programs, provided the views of young people themselves are incorporated in the program content.

(JMIR Mhealth Uhealth 2019;7(10):e11205) doi:10.2196/11205

KEYWORDS

weight gain; young adults; obesity; public health; focus groups; mobile apps; mHealth

Introduction

Background

Obesity is one of the biggest public health challenges of the 21st century. At least 650 million adults are currently obese worldwide [1], and effectiveness of obesity treatments is rather modest [2]. Effective prevention methods are needed to halt the obesity epidemic.

Obesity develops with age and can only be prevented before it has developed. The transition from adolescence to young adulthood is a critical life period with rapid weight gain [3-5]. We currently have the largest ever generation of adolescents and young people and the most vulnerable to obesity and secondary noncommunicable diseases in human history [6]. Effective promotion of sustainable health behaviors to prevent obesity must specifically target people where primary prevention is still possible. At age 18 years, the prevalence of body mass index (BMI)>30 kg/m^2 (5.8%) and BMI>25 kg/m^2 (22.8%), are still relatively low, which more than double by age 35 years [7]. Body weights rise most rapidly between ages 13 to 24 years internationally [8-14], with steeper weight gain trajectories most likely to reach BMI>30 kg/m^2, now affecting 40% by age 65 years in the United Kingdom [15].

Adolescence and young adulthood present vital opportunities for intervention, with multiple lifestyle changes for emerging young adults—self-determining, financially independent, outside parental controls [16]. Social interactions often revolve around heavily marketed foods and drinks, tending to promote greater consumption [17]. Young people tend to be relatively neglected by conventional health promotion, considered *hard to reach* through lack of interface with health information dissemination or engagement with health professionals [18]. The internet has become the dominant mode for communication, information exchange, and increasingly for health care delivery. Thus, electronic health (electronic technology in health care) and mobile health (mHealth; mobile technology in health care) offer numerous advantages above traditional methods [19]. Although conventional didactic methods lack reach, innovative electronic learning and support can prevent unwanted weight gain in young adults [20]. Transferring existing online resources with proven effectiveness for weight gain prevention into mobile apps could reach a much greater proportion of the at-risk young population. Near-universal mobile phone ownership, projected to reach 5 billion by 2019, has been matched by increasing availability and access to mHealth apps, mostly for weight loss and fitness [21]. Mobile app downloads reached 149.3 billion in 2016 [22]. Over 25,000 apps are currently aimed solely at weight management [23], but few (0.17%) report having incorporated experts' and users' inputs during development. Even fewer have been tested for effectiveness, mostly with poor results [24,25]. Efforts to treat established obesity in young people have been rather ineffective [26,27], and there are no apps specifically tailored to offer weight gain prevention for young people.

Objectives

This study used newer information technology (IT) methods to explore the views of young people from several countries on weight gain and its prevention and their opinions and preferences to inform development of weight gain prevention apps tailored to the needs of young people.

Methods

Overview

A *mixed methods* study design was used, with an online survey followed by online asynchronous focus groups [28]. This approach integrates quantitative and qualitative data within a single investigation, to permit a more complete and synergistic utilization of data than separate quantitative and qualitative data collection and analyses [29,30]. The study was approved by the Institutional Ethical Review Board of the Catholic University of Louvain.

Questionnaire Design and Recruitment

The recruitment questionnaire comprised 18 questions: 2 multiple choice questions (country of residence, current occupation), 7 dichotomous questions (parents' academic degree, parents' smoking, participants' smoking, concern about weight gain, interest in a quiz to estimate risk of weight gain, current use of mobile apps, interest in an app specifically designed by researchers on weight gain prevention), 8 demographic questions (gender, age range 13-24 years, postcode, current weight, current height, current waist size, dress size, email address), and 1 open-ended question (if you are using lifestyle mobile apps, which ones are you currently using; Multimedia Appendix 1).

The questionnaire was developed following an iterative process, through discussions with all coauthors based in the different countries participating in this study, to ensure that the questions were culturally appropriate. A draft questionnaire was trialed among colleagues first, and after necessary edits, the final questionnaire was trialed with a small group of young people to ensure all functions worked in the online environment and that the wording of the questions was appropriate for this age group, especially for the countries where English is not the official first language.

This was a closed survey, available only to young people aged between 13 to 24 years and residing in 1 of the 6 countries (Belgium, Singapore, New Zealand, Finland, UK, and Greece). Young people were recruited via an online advertisement service (Facebook). The advertisements appeared in a pop-up format, in the profiles of those who declared their age to be between 13 to 24 years and residing in 1 of 6 countries where English is either an official language (United Kingdom, New Zealand, and Singapore) or widely spoken and understood among young people (Greece, Belgium, and Finland). The advertisement invited young people to help guide researchers in building a mobile app, with a link to the online recruitment questionnaire. Participants were not required to have any particular interest in weight management or health, and no financial or other incentives were offered. The recruitment questionnaire was available in 2 screens in total; the first containing the subject information sheet with information on the study and the second containing the questions. The collected data were managed by SurveyMonkey, where they were automatically recorded and

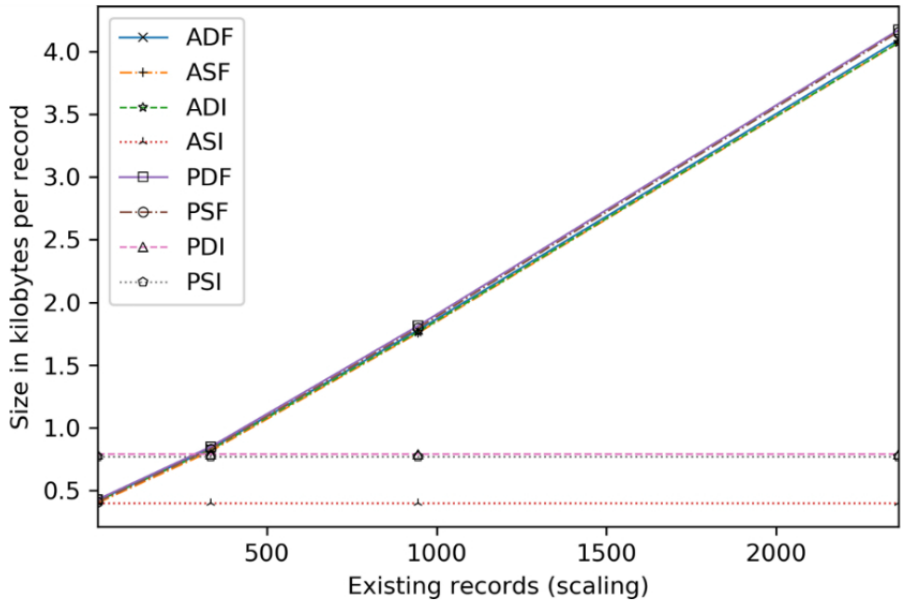
saved. When the advertisement and recruitment closed, data were downloaded and transferred to a university computer, which is encrypted to ensure data protection. Internet protocol addresses were recorded automatically when participants completed the online survey and checked to ensure no multiple entries were submitted by the same participant.

A subject information sheet with information on the study and the principal investigator and contact details to express any concerns they might have on the questionnaire was incorporated in the first page of the questionnaire. People following the link could read the information and decide if they wanted to proceed with the questionnaire. No data were stored for incomplete questionnaires. Respondents were assured anonymity but also invited to provide their email address if they wished to participate in an online focus group in which some of the questions could be discussed further. Those who volunteered to participate in the focus groups were first separated by age group into *adolescents* (13-17 years) or *young adults* (18-24 years) and then block-randomized into 8 groups by gender and country, so that these characteristics were uniformly distributed across groups (Figure 1). The focus groups were conducted on an online platform in password-protected Web rooms (ProBoards). The focus groups were asynchronous, open 24 hours/day for 2 weeks to accommodate participants from all different time zones and life/school/work schedules, allowing more views to be documented.

The focus group moderator (CKN) posted information on the study, ground rules on group etiquette, confidentiality, ethical considerations and consent, and a discussion guide in every Web room. A discussion guide (Multimedia Appendix 2) was created to guide the focus groups, with open-ended questions and prompts to introduce preselected topics: (1) current use of lifestyle apps, (2) factors motivating downloading and retaining apps, (3) use of social media, (4) concerns about weight gain, and (5) concerns about environmental and ethical issues around food and drink production and distribution. Topics 3 and 4 (Multimedia Appendix 2) were included based on the results of a previous study in young adults that found 2 different behavioral change theories: the *rational* and the *stealth* theories to be effective in preventing the usual weight gain in this population [20]. The rational choice theory by Simon (1955) [31] states that humans, provided they have the right information, will make a rational choice [31], whereas the stealth theory by Robinson (2010) [32] states that humans can change a behavior if they find a behavior that is motivating in itself [32].

Online focus groups have advantages and disadvantages compared with conventional face-to-face focus groups. They can involve large numbers; an anonymity may permit more open answers: the moderator emphasized that all responses were welcomed, that none would be considered right or wrong, but that they were monitored for any inappropriate content or behavior. Being unable to share facial expressions, which contribute to face-to-face focus groups, participants were encouraged to use emoticons in their postings which were used in the analysis.

Figure 1. Study procedures for participants completing the survey and those selected to participate in the focus groups. FG: focus group.



Statistical Analysis

Descriptive statistics used SPSS 24 (IBM). Differences between groups were tested with the MedCalc statistical software (MedCalc). Weight, height, and BMI centiles for adolescents (aged 13-18 years) were calculated according to the International Obesity Task Force [33].

Qualitative Analysis

Transcribed data were transferred to software (NVivo, version 11, QSR International) for analysis. Thematic analysis was chosen as the most appropriate analytical method through its ability to identify patterned meanings across a dataset and responses [34]. Coding and analysis were carried out using as described by Braun and Clark [35]. Analysis of transcripts started with multiple readings by the principal researcher (CKN)

to identify keywords and phrases. Relevant words and phrases identified were coded, and data were categorized describing main issues reported by participants.

All initial codes relevant to the discussion guide and research question were grouped into themes by combining similar codes. Codes and themes were triangulated between the first and last authors to enhance the validity of the coding phrases, and all authors reviewed the selective codes, themes, and exemplar quotes. *Main themes* emerged from relevant combinations of named themes, each accompanied by a detailed analysis, to explain the salient sections of data collected from the focus groups.

Reflexivity Statement

Reflexivity relates to sensitivity to the ways in which the researcher and the research process may shape the data collected, including the role of prior assumptions and experience. This study was carried out online; hence, the researcher had minimal contact with the participants during recruitment and conducting of the focus groups. The participants could enter the Web room at any time and from any place when they felt comfortable to do so. Our study was designed to elicit contributions from a broad range of participants. During the focus group process, no individual's views were preferred over those of others. A goal of data analysis was the identification of common themes that emerged from comparison across the focus groups. However, equal importance was attached to the analysis focusing on the details of individuals' reports relating to specific views and experiences as well as those of the groups as a whole.

Results

Overview

Results are reported according to the Checklist for Reporting Results of Internet E-Surveys [36].

Over an 84-hour period after posting the Facebook advertisement, 2285 young people aged 13 to 24 years (610 *adolescents* aged 13-17 years and 1675 *young adults* aged 18-24 years) responded to the advertisement and completed the survey. Demographic information for those completing the survey is presented in Tables 1 and 2.

Young people (1645/2285, 71.99%) reported being concerned about weight gain for themselves or friends, with more young adults (aged 18-24 years) being concerned than adolescents (aged 13-17 years; χ^2_1=18.6, P<.001). There were no differences between countries among young adults for expressing concern about unwanted weight gain, but there was a greater range for adolescents, for example, between Finland (55.4%) and Greece (76.2%). A similar percentage of young people (1657/2285, 73.30%) reported being interested in a weight gain prevention app, again with significantly more young adults being interested in such an app than adolescents (χ^2_1=11.4, P=.001). Approximately half of the young adults (49.7%, 833/1676) and one-third of the adolescents (31.0%, 189/610) reported currently using a lifestyle mobile app with significantly more young adults currently using 1 (χ^2_1=–172.4, P<.001). Among young people, there was no difference in the BMI of those who reported using lifestyle apps and those who did not (21.9 kg/m^2-22.0 kg/m^2). Reasons mentioned by participants for not using apps were as follows: time-consuming, not reliable information/source, and too many advertisements. Similarly, for adolescents, there was no significant difference in BMI centiles of those who reported using a lifestyle app and those who did not. In adolescents, reported use of lifestyle apps increased with age, from 27.6% at age 13 years to 40.8% at age 17 years, (χ^2_1=23.5, P<.001).

Table 1. Demographic characteristics of adolescents completing the survey and those selected to participate in the 6 online focus groups.

Demographic characteristics	Survey (n=610)	Focus group 1 (n=64)	Focus group 2 (n=62)	Focus group 3 (n=64)	Focus group 4 (n=62)	Focus group 5 (n=62)	Focus group 6 (n=62)
Age (years), mean (SD)	15.3 (1.4)	15.0 (1.4)	15.0 (1.4)	15.0 (1.4)	15.0 (1.4)	15.0 (1.4)	15.0 (1.4)
Gender (female), n (%)	429 (70.3)	36 (57)	35 (57)	36 (57)	40 (65)	42 (67)	40 (65)
Weight (kg), mean (SD)	59.7 (8.0)	58.2 (7.9)	60.9 (8.3)	57.9 (7.0)	59.5 (9.8)	60.7 (8.2)	59.03 (6.6)
Height (m), mean (SD)	1.64 (0.07)	1.64 (0.07)	1.64 (0.07)	1.63 (0.06)	1.63 (0.07)	1.64 (0.06)	1.64 (0.07)
BMI[a] (kg/m^2), mean (SD)	22.2 (8.0)	21.4 (1.9)	22.4 (2.07)	21.6 (1.8)	22.02 (2.4)	22.4 (2.2)	21.9 (1.8)
Country, n (%)							
Greece	97 (15.9)	11 (17)	11 (17)	10 (15)	10 (15)	11 (17)	10 (16)
Singapore	93 (15.2)	11 (17)	11 (17)	11 (17)	11 (17)	9 (15)	8 (14)
Finland	74 (12.1)	11 (17)	11 (17)	10 (15)	10 (15)	9 (15)	10 (16)
United Kingdom	163 (26.7)	11 (17)	11 (17)	11 (17)	11 (17)	11 (17)	12 (18)
Belgium	105 (17.2)	11 (17)	11 (17)	11 (17)	11 (17)	11 (17)	12 (18)
New Zealand	78 (12.8)	9 (13)	7 (13)	11 (17)	9 (15)	11 (17)	10 (16)
Parents' higher education degree (yes), n (%)	433 (71.0)	46 (72)	45 (72)	43 (67)	38 (61)	45 (72)	52 (83)
Parents smoking (yes), n (%)	152 (24.9)	18 (27)	17 (27)	13 (20)	20 (31)	17 (27)	18 (29)
Current use of lifestyle apps (yes), n (%)	189 (31.0)	23 (36)	22 (36)	23 (36)	17 (28)	20 (32)	29 (47)

[a]BMI: body mass index.

Table 2. Demographic characteristics of young adults completing the survey and those selected to participate in the focus groups.

Demographic characteristics	Survey (n=1675)	Focus group 1 (n=72)	Focus group 2 (n=73)	Focus group 3 (n=71)	Focus group 4 (n=70)	Focus group 5 (n=73)	Focus group 6 (n=72)
Age (years), mean (SD)	19.7 (1.7)	20.7(1.9)	20.7(1.9)	20.8 (1.9)	20.7 (1.9)	20.7 (1.9)	20.6 (1.9)
Gender (female), n (%)	1170 (69.9)	43 (60)	42 (58)	40 (56)	39 (55)	43 (60)	42 (58)
Weight (kg), mean (SD)	63.8 (11.3)	64.5 (10.2)	63.2 (10.3)	63.5 (10.5)	64.3 (13.5)	62.6 (10.4)	64.2 (10.8)
Height (m), mean (SD)	1.80 (4.3)	1.72 (0.08)	1.71 (0.08)	1.70 (0.09)	1.70 (0.1)	1.70 (0.1)	1.70 (0.07)
BMI[a] (kg/m^2, mean (SD)	21.9 (3.2)	21.6 (2.6)	21.5 (2.8)	21.7 (2.9)	22.0 (3.9)	21.6 92.4)	21.9 (3.0)
Country, n (%)							
Greece	316 (18.86)	10 (14)	11 (15)	11 (16)	11 (16)	12 (16)	12 (16)
Singapore	270 (16.11)	10 (14)	10 (14)	9 (14)	10 (15)	10 (14)	11 (15)
Finland	211 (12.59)	13 (17)	14 (18)	13 (17)	11 (16)	12 (16)	12 (16)
United Kingdom	420 (25.07)	14 (18)	14 (18)	14 (18)	15 (19)	14 (18)	14 (19)
Belgium	219 (13.07)	12 (16)	11 (15)	11 (16)	10 (15)	13 (17)	11 (15)
New Zealand	239 (14.26)	13 (17)	13 (17)	13 (17)	13 (17)	12 (16)	12 (16)
Occupation, n (%)							
In higher education	767 (45.79)	31 (43)	41 (56)	32 (45)	28 (39)	34 (47)	39 (54)
Employed	670 (40.00)	31 (43)	26 (35)	28 (40)	32 (45)	28 (37)	28 (38)
Looking for job/other	238 (14.20)	10 (13)	6 (7)	11 (14)	10 (15)	11 (14)	5 (6)
Parents' higher education degree (yes), n (%)	1280 (76.41)	59 (82)	55 (75)	53 (75)	48 (68)	56 (77)	57 (79)
Parents' smoking (yes), n (%)	390 (23.28)	17 (23)	12 (17)	12.2 (17)	14 (20)	16 (21)	19 (26)
Smoking (yes), n (%)	273 (16.29)	15 (20)	13 (18)	10 (14)	12 (17)	12 (16)	14.2 (18)
Current use of lifestyle apps (yes), n (%)	832 (49.67)	34 (47)	35 (48)	44 (62)	40 (57)	31.5 (43)	41 (56)

[a]BMI: body mass index.

From the 2285-young people who completed the survey, 807 young people (376 adolescents aged 13-17 years and 431 young adults aged 18-24 years) offered and were contacted to participate in the focus groups, stratified by country and age. A total of 12 focus group sessions, held in total at the start of 2017, were completed by 719 young people (completion rate 89%). Detailed demographic characteristics of focus group participants are presented in Tables 1 and 2. Due to the large number of participants, some quantitative analysis was possible. The mean number of individual entries into each focus group for the adolescents was 98 (SD 6.0) and for the young adults 121 (SD 12.0). The moderator tried to observe rather than participate to allow participants to freely express themselves. The moderator checked the Web rooms for any inappropriate content every 12 hours and interacted individually with a total of 23 participants. Moreover, because of the nature of the asynchronous groups and participants being based in 6 countries, all but 2 (Belgium and Finland) in different time zones, participants and moderator logged in at different times and they were not online concurrently.

Main Themes

Main themes emerging from both adolescents' and young adults' focus groups are identified and presented with a general explanation of findings and exemplar quotes. Some of the main themes emerging, unsurprisingly, matched the topics which had been introduced into the focus groups.

Theme 1: Feelings Toward Personal Weight

Young people in all countries reported that weight changes are of concern for them (568/719, 78.9% of participants). Although most participants were neither overweight nor obese, they considered body weight an important indicator of health and fitness. To the adolescent participants, weight changes were considered an important part of the *growing-up* process, but they questioned what is considered *normal* weight gain. Adolescents also expressed concerns about the best way to voice their issues around body weight, diet, and physical activity to parents and friends, with difficulties making lifestyle choices free from families' or friends' influences and used emoticons expressing negative feelings, for example, sad face, face crying, for unwanted weight gain:

> *I think I am really fat because previous cloths don't fit me. And I think I am going to get health problems according to my weight. I do exercise regularly. It would be awesome to have some info on what is normal weight.* [Male, 14 years, Finland]

> *I'm one of the tallest girls in the year, and scared one of the heaviest as well, people keep saying it's because*

> *of your height but is height the only reason or should I be doing more things? I don't think I eat that much and I already do gym 5 hours a week but is this not enough maybe?* [Female, 13 years, New Zealand]

Young adults discussed how weight changes related to taking full responsibility for their lifestyles, while the other significant changes in their lives were occurring. Factors identified for weight increase were types of food available, food costs, and promotion of unhealthy foods such as crisps and chocolate. Increased alcohol consumption was another reason identified for weight gain, particularly by young adults in the United Kingdom and New Zealand, invoking use of emoticons, mostly of sad face:

> *I put on a lot of weight. Clothes do not fit me. I tried with my friends to lose weight with things like going for a run but then we would go out drinking so...didn't lose any weight in the end.* [Male, 22 years, New Zealand]

Theme 2: Perceptions of Lifestyle Apps and Desired Content of a Weight-Gain Prevention App

Most participants (546/719, 75.9%) said that they would be interested in app access to electronic information on weight gain prevention. All young people reported being already exposed to abundant information on different diets, *super foods*, and physical activity regimes through websites, magazines, or articles shared through social media. This plethora of information, often conflicting, was difficult to process, and they would welcome advice and information from a source they could trust. Young people again used emoticons to illustrate these feelings, of face confused, face spinning, and sad face. Having access to trustworthy and evidence-based information from what they consider a reliable source was very commonly mentioned (n=489, 68.0% of participants). They mistrust and cannot connect with the general guidelines provided by government agencies but think that information coming from independent researchers and health care professionals would be trustworthy:

> *You read all this news like eat chocolate because it helps with depression one day and then the other don't because it has a lot of sugar. Not sure what to believe. Someone like a doctor needs to filter this information.* [Female, 14 years, Greece]

> *There is so much info on diets, nutrition, exercise out there. It feels like constantly being bombarded by so many things and in school you are being told something else and at home something else. Those people doing research seem to know things better. Maybe they could give monthly updates on these topics with accurate information.* [Female, 16 years, Belgium]

On the basis of previous research on online resources for young people to prevent unwanted weight gain, participants were asked to discuss what type of information should be included in a weight gain prevention app. Examples and prompts from the previous study were provided [20]. Most identified diet as the most important factor for weight changes. They wanted information on what is good to eat without promoting excessive weight gain, particularly snacks, and what are *normal* weight gain trajectories.

When asked about environmental and commercial issues around food, they expressed concern over whether chemical or antibiotics had been used and whether people are forced to work in certain food production sectors. These issues/concerns were reported mostly from young people older than 16 years. Participants also discussed wider geopolitical aspects of food practices, especially how corporations producing and handling food affect consumers, society, and the environment including climate change. Adolescents in all countries reported having school lessons on environmental issues, associating certain food habits with environmental issues and current social movements. They wanted better information over how food choices affect the planet and felt it should be online to better inform decisions and actions for more people. They proposed either joining existing social movements, or creating new ones, according to local needs and characteristics:

> *I read about the food miles. It is crazy when we have tomatoes or apples produced here to buy from other countries. I don't understand why. This can't be good for anyone.* [Male, 17 years, Greece]

> *I read in the news often about people working in the fish boats like slaves. I don't want to consume the product of this trade.* [Male, 19 years, New Zealand]

> *There is a community garden close to my home, and sometimes I go with my dad there. I think it is a great idea to use this land to get local food. A lot better than buying food from the supermarket that travelled from the other side of the world.* [Female, 18 years, United Kingdom]

Theme 3: Social Media Apps, Lifestyle Apps, and Motivation for Downloading and Retaining

Young people reported using some kind of app every day, mostly those linked to social media and instant communication (719/719, 100% of participants), the most popular cited being Facebook, Instagram, Twitter, and Snapchat. This finding was similar across different countries. Young people reported using Facebook mostly for its messaging service or in closed groups for school work. Instagram is very popular because it is considered more private and less commercialized than other social media:

> *Facebook is kind of dead. Everybody used to have it and most of us will still do but not really for posting anything.* [Male, 16 years, Belgium]

> *Instagram hasn't been flooded with the older generation yet (not everyone has an Instagram), it's "hip" and "cool." There are no links.* [Female, 17 years, New Zealand]

Young people reported using Twitter for following news and people they are interested in but not for posting new content of their own as they feel it is less private. Snapchat was felt to carry a lot less social pressure, allowing young people to be more authentic, and to be both addictive and liberating. They felt closed groups on Facebook were most suitable for sharing

longer pieces of information on weight gain prevention, and Snapchat or Twitter for short messages.

Young people were not aware of the entire range of lifestyle apps (diet, physical activity) currently available on the market. When they reported using lifestyle apps, these were either 1 of the most popular apps advertised in app shops or had been recommended by friends or family. Young people reported using those for weight management and particularly for increasing/monitoring their physical activity. Some reported using the default lifestyle apps incorporated into their mobile phone device. Poor retention was frequent for lifestyle apps: the main reasons for discontinuing were taking too much time to use them and find the information they wanted and receiving too frequent and too general reminders, which participants found annoying:

> *I thought I wanted to have a calorie input thing, so that I can put how many calories I had that day but it took too much time and got bored of it soon.* [Female, 23 years, Singapore]

Finding the right motivation to exercise and eat healthily is also important for young adults:

> *I'm quite lazy and need motivation to exercise and be fit. When living at home, my mum always reminded me of exercising and cooked nice meals. Would really love if someone would message me and maybe motivate each other online?* [Male, 19 years, Belgium]

Theme 4: Data Safety and Usage, Confidentiality

During the discussion on what technical features young people would like in a lifestyle app, the issues of safety, data usage, and confidentiality emerged most frequently. They were particularly concerned about releasing personal details or allowing apps to have access to personal data or other accounts (emails, Facebook, etc) on mobile phones. They were aware of their *digital footprints*. They did not wish to post traceable personal data or opinions, as they may change their mind later, and such posts might disadvantage them in the future. Online reputation and identity seem to be a very important discussion point among young people (438, 61.0% of participants). However, they would value some form of support to encourage them to continue using an app, for example, *chat bots*, and they were happy to engage with their peers via anonymous online chats or direct SMS (short message service) text messaging:

> *Your tweets are easily searchable on Twitter which is ok but not good if you want to be yourself.* [Male, 20 years, United Kingdom]

> *There are some apps that collect a lot of your personal data. They ask for access to your contacts, photos etc. I have never thought about it previously but why an app needs to have access to my personal stuff?* [Male, 18 years, New Zealand]

> *I had apps downloaded that had all of these ads on random things, some quite inappropriate. I don't want to view all these adds.* [Female, 16 years, Finland]

Data usage with mobile apps was another important issue. Most young people are on restricted budgets with limited mobile data. Despite the high internet coverage, some young people reported not having internet at home because of cost. It is of great importance to them to avoid draining their battery or mobile data:

> *Some apps just take up all my data. Maybe have something like a light version like the one Facebook has, the messenger lite.* [Female, 17 years, Greece]

Discussion

Principal Findings

Our results show that young people from 6 different countries are not only concerned about weight changes and diet but also about the relationships with the environment they live in. They would like independent advice to help them negotiate the environment to control weight changes, with lifestyle choices, which also help protect the environment. They are clear that guidance should be in a format that is familiar to them and easily accessible, that is, through a mobile device as long as their concerns are being addressed.

To the best of our knowledge, this is the first study to be conducted among young people at a multicounty level on the potential of mobile apps for weight gain prevention. The mixed online methods used made this possible, bringing young people from 6 countries to the same *room* to discuss topics of common interest. It offered some advantages including the opportunity for participants to provide candid honest, opinions from a position of anonymity. Online focus groups permit many more participants to join and avoid the effort and cost of transcribing recorded discussions, but the resulting discussions may be narrower than in conventional one-to-one focus groups where participants may be more inclined to pick up collateral or tangential issues, and there is a greater opportunity to identify cross-cutting themes. The electronic discussion among young people from different countries set up a stimulus-response reaction that revealed young people's preferences, concerns, and barriers to using technology for managing their lifestyle. The participants, from 6 countries, reported experiencing our globalized world in a remarkably similar way, sharing common concerns and preferences. Although commonly concerned about unwanted weight gain and welcoming advice on prevention, they are concerned to have reliable information sources and for their data and identities to be protected. Younger people are reliant on mobile communications. They are well adapted to the wider digital environment and also well aware of its potential perils and how their online activity might affect their future. As experienced users, they want end products to meet closely their requirements and needs and not waste time or money.

The main topics/themes incorporated in the design of our questionnaire, and in leading the focus group discussions, were chosen because they are already established as frequent areas of concern to young people in other contexts. Other topics, such as online safety, data usage, and professional and trustworthy advice arose in the focus group discussions. It was important for this study to confirm that the main known issues apply to the subset of young people who have body weight concerns and

who may be attracted to use apps directed at preventing weight gain. These individuals may have subtly different attitudes toward decision making, health risks, and risks in general, and they are known to be tempted to spend large amounts of money online for products without evidence for value. The results of the questionnaires tend to confirm that young people do indeed have similar areas of concerns around online/app approaches directed at body weight control.

Comparison With Prior Work

There is growing interest in Web-based approaches for health under the current World Health Organization Research Agenda, with a view to develop *Health in your Pocket* interventions [37]. The European Commission echoes this with calls for a digital single market [38]. Obesity has become a global public health issue, affecting both developed and developing countries. Solutions that have the potential to be scaled up to reach large populations are therefore vital.

Mobile apps seem to have the potential to increase motivation to eat a more healthful diet [39], but limited research is available about the use of mobile apps for health promotion for adolescents [40] and young people. The digitalized world, which younger generations are in tune with, offers ways to implement solutions that are affordable and have global reach [41]. By the year 2020, an entire generation will have grown up in a fully digital world, with computers, internet, mobile phones, and social networking all being second nature.

Data Safety/Reliable Independent Information

About one-third of adolescents and half the young adults reported currently using lifestyle apps. This agrees with a previous study where lifestyle apps were reported to be used by 27.6% of adolescents [42]. Those currently using lifestyle apps mostly use either the most popular apps available in app stores or those recommended by close friends, family, or a trustworthy source rather than government agencies. Young people are especially concerned about their data confidentiality and online safety, with good reason. In a survey conducted in the United Kingdom for Safe Internet Day in 2016, it was showed that in the previous year, 82% of teenagers saw hateful content unintentionally on the internet [43]. Privacy policies incorporated into mobile apps targeting young people rarely adapt their policies to the language style or reading level of young people [44].

Environmental Protection/Commercial Exploitation/Ethics

Finding the right motivation to engage with weight gain prevention tools is critical for effectiveness. A recent study found that introducing *rewards* provided good motivation to use a dietary assessment app [45]. Another study using young adults' focus groups found that the use of social media and mobile gaming is an acceptable method for increasing vegetable consumption [46]. Aligning messages to the topical concerns of young people may strengthen the motivation for behavior change: a recent survey conducted by United Nations Children's Fund and Eurochild found that 41% of the 15,000 young people participating were worried about climate change [47]. This is in agreement with our results, showing that young people are worried about the effect that their food choices may have on the planet and on people involved in food production.

These sentiments suggest promoting *Food Citizenship*: "the practice of engaging in food-related behaviors that support, rather than threaten, the development of a democratic, socially and economically just, and environmentally sustainable food system" [48].

Mobile services and devices have had a sudden boom that allowed very little time for involving the end users in the product designs. Many lifestyle apps are available, but few are evidence-based or had input from the end users in the design process, which may limit success and effectiveness [49,50]. Involving young people in the design process, piloting, and marketing creates a sense of *ownership*, promoting engagement and motivation for using the app [51]. Given the enormous number of current unreliable and non-evidence-based lifestyle apps in the market, it would be reasonable to try first to transfer the already tested and validated resources into mobile formats. Our study provides evidence that young people do want tailored apps, and the insights from online focus groups can inform appropriate app development to meet the needs and preferences of young people as reported by them. Specifically, the process for creating a user-friendly and engaging mobile app on weight gain prevention needs to address safety, cost, online reputation, and the content desired by young people.

Strengths and Limitations

Although the issues appear very similar in the 6 countries studied, there are limitations. In this study, only English language was used for the online focus groups. The countries were selected as those where large proportions of young people speak English at a high standard, but the participants may have been individuals with relatively high educational attainment, and their views are not necessarily reflective of non-English speakers in those countries. To take account of this issue, we offered the recruitment advertisement in both English and the local languages. The high consistency in results between countries where English is usually the first language across social and educational classes (United Kingdom, Singapore, New Zealand) and those where English is usually a second language (Belgium, Greece, Finland) suggests that the results may be widely generalizable. We did not include developing countries where socially determined obesity risks and diets are very different, and mobile phone ownership is lower. This study used a relatively new recruitment method for research, with advertisements posted on social media. This ensured maximum visibility across the participating countries and unified adoption with the same approach in all countries. The online method allowed for far more young people to participate than traditional focus groups. The advertisements did not mention weight or health to reduce selection bias and to attract a broad range of participants. Both recruitment and conducting methods were low-cost, convenient, and confidential, suited to reaching young people and for addressing sensitive topics such as body weight. There was no requirement for labor-intensive transcription with online focus groups.

The mixed methods design used in this study also helped with gauging this new topic in a more efficient and comprehensive way, by combing qualitative and quantitative data.

All focus group methods engage with subsectors of the target population that are interested in the topic and have time available. The online approach allows a wide section to participate, in their own time, but does introduce some limitations. The lack of face-to-face interaction and visual cues to influence discussions has been mentioned, but freedom from shyness allows participants to exchange views more freely and avoids intimidation. Furthermore, participants used emoticons to express their feelings. Young people are especially familiar with the use of emoticons, and they feel that emoticons can clarify meaning to the text [52]. Our focus groups were mixed in gender and stratified by age and country. Some researchers have preferred gender-specific focus groups for fear of males dominating the discussion—the *peacock effect* [53]. This effect is less likely in the online environment where participants could log in and express views at any time without threat from others. Online focus groups inevitably have some different characteristics from conventional face-to-face focus groups, being akin to moderated discussion fora, or blogs, and whether they generate the same research information needs separate investigation.

Conclusions

Young people are commonly, and consistently across 6 countries, concerned about weight gain and obesity. Applying no industry evidence-based IT programs to guide young people toward preventing weight gain would be well received, provided that the views of young people themselves are incorporated in the program content and app design.

Acknowledgments

This research was part-funded by a Marie-Curie-Move in Louvain Fellowship awarded to CKN from the European Union.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Recruitment questionnaire.
[PDF File (Adobe PDF File), 70 KB - mhealth_v7i10e11205_app1.pdf]

Multimedia Appendix 2
Discussion guide for the groups.
[PDF File (Adobe PDF File), 364 KB - mhealth_v7i10e11205_app2.pdf]

References

1. World Health Organization. 2018. Obesity and Overweight: Key Facts URL: http://www.who.int/mediacentre/factsheets/fs311/en/ [accessed 2018-08-31] [WebCite Cache ID 725VOSzDN]
2. Mauro M, Taylor V, Wharton S, Sharma AM. Barriers to obesity treatment. Eur J Intern Med 2008 May;19(3):173-180. [doi: 10.1016/j.ejim.2007.09.011] [Medline: 18395160]
3. Dietz WH. Overweight in childhood and adolescence. N Engl J Med 2004 Feb 26;350(9):855-857. [doi: 10.1056/NEJMp048008] [Medline: 14985480]
4. Lawlor DA, Chaturvedi N. Treatment and prevention of obesity--are there critical periods for intervention? Int J Epidemiol 2006 Feb;35(1):3-9. [doi: 10.1093/ije/dyi309] [Medline: 16396899]
5. Alberga AS, Sigal RJ, Goldfield G, Prud'homme D, Kenny GP. Overweight and obese teenagers: why is adolescence a critical period? Pediatr Obes 2012 Aug;7(4):261-273. [doi: 10.1111/j.2047-6310.2011.00046.x] [Medline: 22461384]
6. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jun 11;387(10036):2423-2478 [FREE Full text] [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
7. European Commission. 2017. Overweight and Obesity - BMI Statistics URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Overweight_and_obesity_-_BMI_statistics [accessed 2018-08-31] [WebCite Cache ID 725VlHNFm]
8. Gordon-Larsen P, Adair LS, Nelson MC, Popkin BM. Five-year obesity incidence in the transition period between adolescence and adulthood: the National Longitudinal Study of Adolescent Health. Am J Clin Nutr 2004 Sep;80(3):569-575. [doi: 10.1093/ajcn/80.3.569] [Medline: 15321794]
9. Neumark-Sztainer D, Wall M, Eisenberg ME, Story M, Hannan PJ. Overweight status and weight control behaviors in adolescents: longitudinal and secular trends from 1999 to 2004. Prev Med 2006 Jul;43(1):52-59. [doi: 10.1016/j.ypmed.2006.03.014] [Medline: 16697035]
10. Lewis CE, Jacobs DR, McCreath H, Kiefe CI, Schreiner PJ, Smith DE, et al. Weight gain continues in the 1990s: 10-year trends in weight and overweight from the CARDIA study. Coronary artery risk development in young adults. Am J Epidemiol 2000 Jun 15;151(12):1172-1181. [doi: 10.1093/oxfordjournals.aje.a010167] [Medline: 10905529]

11. Vella-Zarb RA, Elgar FJ. The 'freshman 5': a meta-analysis of weight gain in the freshman year of college. J Am Coll Health 2009;58(2):161-166. [doi: 10.1080/07448480903221392] [Medline: 19892653]
12. Nikolaou CK, Hankey CR, Lean ME. Weight changes in young adults: a mixed-methods study. Int J Obes (Lond) 2015 Mar;39(3):508-513. [doi: 10.1038/ijo.2014.160] [Medline: 25152239]
13. Wardle J, Haase AM, Steptoe A. Body image and weight control in young adults: international comparisons in university students from 22 countries. Int J Obes (Lond) 2006 Apr;30(4):644-651. [doi: 10.1038/sj.ijo.0803050] [Medline: 16151414]
14. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in diet and lifestyle and long-term weight gain in women and men. N Engl J Med 2011 Jun 23;364(25):2392-2404 [FREE Full text] [doi: 10.1056/NEJMoa1014296] [Medline: 21696306]
15. Vlassopoulos A, Combet E, Lean ME. Changing distributions of body size and adiposity with age. Int J Obes (Lond) 2014 Jun;38(6):857-864. [doi: 10.1038/ijo.2013.216] [Medline: 24247373]
16. Lule E, Rosen JE, Singh S, Knowles JC, Behrman JR. Adolescent health programs. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, et al, editors. Disease Control Priorities in Developing Countries: A Summary. Washington, DC: World Bank; 2006:1109-1126.
17. Standing Committee on Childhood Obesity Prevention, Food and Nutrition Board, Institute of Medicine. In: Breiner H, Parker L, Olson S, editors. Challenges and Opportunities for Change in Food Marketing to Children and Youth: Workshop Summary. Washington, DC: National Academies Press; 2013.
18. Montero K, Kelly P. Young People and the Aesthetics of Health Promotion: Beyond Reason, Rationality and Risk. Abingdon, United Kingdom: Routledge; 2016.
19. Kay M, Santos J, Takane M. mHealth: new horizons for health through mobile technologies. World Health Organization 2011;64(7):66-71 [FREE Full text]
20. Nikolaou CK, Hankey CR, Lean ME. Elearning approaches to prevent weight gain in young adults: a randomized controlled study. Obesity (Silver Spring) 2015 Dec;23(12):2377-2384 [FREE Full text] [doi: 10.1002/oby.21237] [Medline: 26538383]
21. European Commission. 2014. Digital Health: mHealth URL: https://ec.europa.eu/digital-single-market/en/mhealth [accessed 2018-08-31] [WebCite Cache ID 725VxHfqt]
22. Clement J. Statistica. 2019. Mobile App Usage - Statistics & Facts URL: https://www.statista.com/topics/1002/mobile-app-usage/ [accessed 2018-08-31] [WebCite Cache ID 725WFvwXK]
23. Nikolaou CK, Lean ME. Mobile applications for obesity and weight management: current market characteristics. Int J Obes (Lond) 2017 Jan;41(1):200-202. [doi: 10.1038/ijo.2016.186] [Medline: 27780974]
24. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. J Am Med Assoc 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
25. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. J Am Med Assoc 2016 Sep 20;316(11):1161-1171 [FREE Full text] [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
26. Svetkey LP, Batch BC, Lin PH, Intille SS, Corsino L, Tyson CC, et al. Cell phone intervention for you (CITY): a randomized, controlled trial of behavioral weight loss intervention for young adults using mobile technology. Obesity (Silver Spring) 2015 Nov;23(11):2133-2141 [FREE Full text] [doi: 10.1002/oby.21226] [Medline: 26530929]
27. Godino JG, Merchant G, Norman GJ, Donohue MC, Marshall SJ, Fowler JH, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755 [FREE Full text] [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
28. Williams S, Clausen MG, Robertson A, Peacock S, McPherson K. Methodological reflections on the use of asynchronous online focus groups in health research. Int J Qual Methods 2012 Sep;11(4):368-383 [FREE Full text] [doi: 10.1177/160940691201100405]
29. Wisdom J, Creswell JW. PCMH Resource Center. 2013. Mixed Methods: Integrating Quantitative and Qualitative Data Collection and Analysis While Studying Patient-Centered Medical Home Models URL: https://pcmh.ahrq.gov/page/mixed-methods-integrating-quantitative-and-qualitative-data-collection-and-analysis-while [accessed 2019-08-07]
30. McKim CA. The value of mixed methods research: a mixed methods study. J Mix Methods Res 2016 Jul 8;11(2):202-222. [doi: 10.1177/1558689815607096]
31. Simon HA. A behavioral model of rational choice. Q J Econ 1955 Feb;69(1):99-118. [doi: 10.2307/1884852]
32. Robinson TN. Stealth interventions for obesity prevention and control: motivating behavior change. In: Dube L, Bechara A, Dagher A, Drewnowski A, LeBel J, James P, et al, editors. Obesity Prevention: The Role of Brain and Society on Individual Behavior. Massachusetts, United States: Academic Press; 2010:319-327.
33. World Obesity Federation. 2017. International Obesity Task Force (IOTF) Body Mass Index Cut-Offs for Thinness, Overweight and Obesity in Children URL: https://www.worldobesity.org/data/cut-points-used/newchildcutoffs [accessed 2019-08-07]
34. Guest GS, MacQueen KM, Namey EE. Applied Thematic Analysis. Thousand Oaks, CA: Sage Publications; 2011.
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

36. 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]
37. van Heerden A, Tomlinson M, Swartz L. Point of care in your pocket: a research agenda for the field of m-health. Bull World Health Organ 2012 May 1;90(5):393-394 [FREE Full text] [doi: 10.2471/BLT.11.099788] [Medline: 22589575]
38. European Commission. 2015. Shaping the Future: Youth and the Digital World URL: https://ec.europa.eu/commission/commissioners/2014-2019/navracsics/announcements/shaping-future-youth-and-digital-world_en [accessed 2019-08-07]
39. West JH, Belvedere LM, Andreasen R, Frandsen C, Hall PC, Crookston BT. Controlling your 'app'etite: how diet and nutrition-related mobile apps lead to behavior change. JMIR Mhealth Uhealth 2017 Jul 10;5(7):e95 [FREE Full text] [doi: 10.2196/mhealth.7410] [Medline: 28694241]
40. Dute DJ, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 5;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
41. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
42. de Cock N, Vangeel J, Lachat C, Beullens K, Vervoort L, Goossens L, et al. Use of fitness and nutrition apps: associations with body mass index, snacking, and drinking habits in adolescents. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e58 [FREE Full text] [doi: 10.2196/mhealth.6005] [Medline: 28442455]
43. European Commission. 2018. Staying Safe Online: A Better Internet for Children URL: https://ec.europa.eu/commission/commissioners/2014-2019/ansip/blog/staying-safe-online-better-internet-children_en [accessed 2018-08-31] [WebCite Cache ID 725o1G2Yw]
44. Das G, Cheung C, Nebeker C, Bietz M, Bloss C. Privacy policies for apps targeted toward youth: descriptive analysis of readability. JMIR Mhealth Uhealth 2018 Jan 4;6(1):e3 [FREE Full text] [doi: 10.2196/mhealth.7626] [Medline: 29301737]
45. Svensson A, Magnusson M, Larsson C. Overcoming barriers: adolescents' experiences using a mobile phone dietary assessment app. JMIR Mhealth Uhealth 2016 Jul 29;4(3):e92 [FREE Full text] [doi: 10.2196/mhealth.5700] [Medline: 27473462]
46. Nour MM, Rouf AS, Allman-Farinelli M. Exploring young adult perspectives on the use of gamification and social media in a smartphone platform for improving vegetable intake. Appetite 2018 Jan 1;120:547-556. [doi: 10.1016/j.appet.2017.10.016] [Medline: 29032184]
47. Eurochild. 2018. The Europe Kids Want: Sharing the Views of Children and Young People Across Europe to Mark World Children's Day URL: https://www.eurochild.org/fileadmin/public/05_Library/Thematic_priorities/05_Child_Participation/Eurochild/Europe_Kids_Want_Brochure.pdf [accessed 2019-02-10] [WebCite Cache ID 765EK6Bdt]
48. Wilkins JL. Eating right here: moving from consumer to food citizen. Agr Hum Values 2005 Sep;22(3):269-273. [doi: 10.1007/s10460-005-6042-4]
49. van de Kar E, den Hengst M. Involving users early on in the design process: closing the gap between mobile information services and their users. Electron Mark 2008 Dec 11;19(1):31-42. [doi: 10.1007/s12525-008-0002-y]
50. Shah SG, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
51. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
52. YouGov | What the world thinks. 2016. Brands Beware: Don't Overdo It With the Emojis URL: https://today.yougov.com/news/2016/11/17/consumers-tired-of-emojis/ [accessed 2018-08-31] [WebCite Cache ID 725WYKT3I]
53. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks, CA: Sage Publications; 2014.

Abbreviations

BMI: body mass index
IT: information technology
mHealth: mobile health

Edited by L Buis; submitted 01.06.18; peer-reviewed by M DeJonckheere, R Newton; comments to author 10.07.18; revised version received 01.09.18; accepted 20.06.19; published 10.10.19.

Please cite as:
Nikolaou CK, Tay Z, Leu J, Rebello SA, Te Morenga L, Van Dam RM, Lean MEJ
Young People's Attitudes and Motivations Toward Social Media and Mobile Apps for Weight Control: Mixed Methods Study
JMIR Mhealth Uhealth 2019;7(10):e11205
URL: https://mhealth.jmir.org/2019/10/e11205
doi:10.2196/11205
PMID:31603431

Original Paper

Design and Development of Smoking Cessation Apps Based on Smokers' and Providers' Perspectives in China: Survey Study

Junfang Xu[1], PhD; Jonathan Bricker[2,3], PhD; Xiaoxing Fu[4], PhD; Chunyan Su[5], PhD; Peicheng Wang[6], MD; Tengfei Qi[7], MA; Feng Cheng[6], PhD

[1]Center for Health Policy Studies, School of Public Health, Zhejiang University School of Medicine, Hangzhou, China

[2]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, United States

[3]Department of Psychology, University of Washington, Seattle, WA, United States

[4]Institute of Anthropology, Renmin University of China, Beijing, China

[5]Media School, University of Chinese Academy of Social Sciences, Beijing, China

[6]Research Center for Public Health, School of Medicine, Tsinghua University, Beijing, China

[7]Department of Sociology, Tsinghua University, Beijing, China

Corresponding Author:
Feng Cheng, PhD
Research Center for Public Health
School of Medicine
Tsinghua University
30 Shuangqing Rd
Beijing, 100084
China
Phone: 86 010 6278 4700
Email: fcheng@mail.tsinghua.edu.cn

Abstract

Background: Although there are more than 60 smartphone apps for smoking cessation in China, many of them do not include the content and features that health care professionals and smokers prefer—which may make them impractical, unengaging, and ineffective. Therefore, we investigated both health care providers' and smokers' preferences for features of future smoking cessation apps.

Objective: This study aimed to investigate Chinese health care providers' and smokers' desired features of a smoking cessation app, with the goal of providing design recommendations for app designers and researchers.

Methods: Both Chinese smokers who own smartphones (n=357) and Chinese health care providers (n=224) responded to a survey collecting data on their sociodemographic characteristics and opinions on the importance of 20 smoking cessation app design features studied in previous US research.

Results: Chinese health care providers expressed strong support of smoking cessation apps on a number of attitude indicators (range 153/224, 68.3% to 204/224, 91.1%). They rated nearly all (18/20) features as very or extremely important (range 52.2%-83.4%) and rated nearly all features (17/20) as more important than the smokers did. More than 60% of smokers rated the following 4 features as very or extremely important: allow sharing the process of smoking cessation with family members and friends (216/319, 67.7%), helping smokers track their progress (such as the amount of smoking per day; 213/319, 66.8%), helping with the side effects of medications and nicotine withdrawal symptoms (201/319, 63.0%), and adapting to ongoing needs and interests of smokers (194/319, 60.8%). Contrary to a similar study of US smokers and health care providers, Chinese smokers and providers rated reputation and ability to communicate with family members and friends as important features, whereas Chinese smokers rated privacy and security as less important.

Conclusions: The design of future smoking cessation and health behavior change apps should consider perspectives of both providers and smokers as well as the role of culture.

(JMIR Mhealth Uhealth 2019;7(10):e12200) doi:10.2196/12200

KEYWORDS

smoking cessation apps; mHealth; design; smartphones; China

Introduction

Background

The harm caused by tobacco use is one of the most serious public health problems in the world. Although smoking has declined in some countries, it remains prevalent in China [1-3]. The 2015 China Adult Tobacco Survey Report [4] estimated the smoking rate among individuals aged 15 years or older to be 27.7%, which means approximately 316 million smokers live in China. The smoking rate among males is 52.1%, with 60.0% of male smokers aged between 45 and 64 years. Currently, smoking kills more than 1 million people each year, and this number is expected to increase to 2 million if current smoking rates continue [5]. The proportion of the smoking population planning to stop smoking within 12 months has increased from 15.4% in 2010 to 17.6% in 2015 [6]. Correspondingly, many smoking cessation strategies are being attempted, such as nicotine replacement therapy, acupuncture, hypnotherapy, and behavioral interventions [7]. However, traditional quit smoking interventions have had very modest success rates and limited population-level reach [8].

There is a technology that offers a potentially impactful opportunity to change that in China: smartphone apps. In 2013, about 40% of the Chinese population, approximately 556 million people, owned smartphones and that increased to almost 90% of Chinese people (1.3 billion) in 2018 [9]. The ubiquity of smartphone use in China and its projected rise over the next 10 years make smartphones a highly accessible platform for delivering quit smoking interventions to millions of Chinese smokers each year. Therefore, 64 smoking cessation apps have been developed in the mobile app market in 2016, providing new ways to assist smokers in quitting smoking [10].

However, Chinese smoking cessation apps have not proven either popular or effective. One challenge is utilization. For example, research shows that 26% of individuals use their downloaded app only once [11]. The user ratings of smoking cessation apps are also low. Results of current studies in the United States [12,13] and a recent study we conducted in China [10] indicate that these apps still lack most elements that are recommended for quitting smoking. Moreover, studies on smoking cessation apps show that many lack the features that health care professionals and smokers prefer, thereby contributing to their low utilization [14,15].

Objectives

To date, 1 study, conducted in the United States, has focused on health providers' and smokers' desired features of smoking cessation apps [16]. The study showed the features of *free or low cost*, *keeps information private*, *matches individual needs and interests*, and *adapts as one's needs and interests change* to be very or extremely important [16]. However, the content and features that render smoking cessation apps engaging and appealing to users in the United States may be considerably different in China, given the differences in cultural values and practices between Chinese and Western cultures [17,18]. Therefore, the aim of this study was to investigate Chinese health care providers' and smokers' desired features of a cessation app, with the goal of providing design recommendations for app designers and researchers.

Methods

Samples

The samples were adult Chinese smokers and health care providers. Regarding the smoker sample, they were volunteers recruited via smoking cessation clinics in Beijing and Web-based advertisements. For those recruited via clinics, all smokers who visited the smoking cessation clinics during May 1, 2016, and June 1, 2016, were invited to participate in our study. In total, 38 eligible smokers with face-to-face interviews were included in our study. For those recruited via Web-based advertisements, we used WeChat (Tencent Inc) and QQ (Tencent Inc), 2 of the most popular multipurpose messaging social media platforms in China. (For the reader's reference, WeChat was downloaded on 94% of China's smartphones, with 806 million active users each month in 2016. The number of active accounts in QQ was 808 million per month in 2014 [19,20].) Interested and eligible smokers were invited to provide Web-based feedback on their preferences for the features of smoking cessation apps. The questionnaires used in the Web-based and in-person interviews were the same. Smokers' inclusion criteria were that they should (1) be aged 18 years or older, (2) own a smartphone, and (3) smoke at least one cigarette every day. The exclusion criteria were that they (1) were aged 18 years or younger, (2) smoked less than daily, (3) could not use apps installed on the smartphone, and (4) had 30% data missing from the questionnaire [21].

Regarding the health care provider sample, they were recruited through convenience sampling and snowball sampling. We started with physicians who worked in pulmonary clinics and specialized in treating nicotine dependence. We then asked them to recommend other related health care providers. In total, 319 smokers and 224 health care providers were included in the data analysis.

Data Sources and Analysis

We collected data using face-to-face interviews with a questionnaire and a Sojump Web-based questionnaire survey software between May 1, 2016, and August 31, 2016, for smokers and March 25, 2019, and May 5, 2019, for health care providers. Sojump is a secured Web-based survey platform that can administer surveys via QQ or WeChat. The survey collected the following information: (1) demographic characteristics and other basic information (Tables 1-3) and (2) a rating of 20 possible smoking cessation app features first described in a previous study of US smokers and providers [16]. The authors deleted 1 feature (ie, program lets you communicate with your personal doctor or health care team) because the vast majority of people in China are not assigned to personal doctors or health care team members. The attitudes and beliefs about smoking cessation apps from both providers' and smokers' perspectives were collected using a 5-anchor rating scale (completely disagree, somewhat disagree, neutral, somewhat agree, and completely agree).

Both health care providers and smokers were asked to rate the importance of each feature using a 4-anchor rating scale (not at all, somewhat, very, and extremely). The features were categorized into 5 domains developed in the previous study of US smokers [16]. The first domain was cost: program is low cost or free. The second domain was reputation: being *research tested*, endorsed by clinical experts, and highly rated by others. The third domain was privacy and security: confidentiality of information, information stored on mobile phone, and information stored in secured *cloud*. The fourth domain was supportive content and user experience: help smokers track their progress, help with the side effects of medications and nicotine withdrawal symptoms, match personal needs and interests, adapt to ongoing needs and interests of smokers, include information about smoking cessation medications, can send out auxiliary or mobility information, include stories about the experiences of quitting smoking from smokers, include videos about quitting smoking, and include games or entertainment. The fifth domain was communication: allow communication between smokers and experts on smoking cessation; allow smokers to communicate with other smokers; allow sharing the process of smoking cessation with family members and friends; and allow smokers to disclose information on microblogs, WeChat, or other social networking sites.

The features were chosen to (1) reflect technology-based strategies for implementing best practice treatment recommendations (eg, addressing use of pharmacotherapy, providing social support, and offering cognitive behavioral–based content), (2) reflect ways to leverage other smartphone capacities to make these programs more engaging (eg, gaming), (3) assess perceived limitations of mobile health (mHealth) tools (eg, security and privacy), or (4) understand other user preferences that may inform future program development (eg, cost and reputation). These features were fully consistent with the US study [16] and thus made it comparable to that study.

Statistical Analysis

Frequencies and percentages were used to describe the opinions or beliefs about smoking cessation apps and the features. We analyzed whether the importance of the rated features differed significantly among subgroups of the sample using the Mann-Whitney *U* and Kruskal-Wallis tests.

Table 1. Demographic characteristics of smokers (N=319).

Variable and category	Statistics
Gender, n (%)	
Male	295 (92.5)
Female	24 (7.5)
Age (years), n (%)	
≤25	83 (26.0)
26-35	127 (39.8)
≥36	109 (34.2)
Occupation, n (%)	
Full time worker	148 (46.4)
Student	40 (12.5)
Other	131 (41.1)
Education level, n (%)	
High school or below	46 (14.4)
Bachelor's or senior college degree	195 (61.1)
Master's degree or above	78 (24.5)
Nationality, n (%)	
Han	283 (88.7)
Other	36 (11.3)
Living region, n (%)	
Metropolis (ie, Beijing, Shanghai, and Guangzhou)	122 (38.2)
Provincial capital[a]	81 (25.4)
Prefecture-level city[b]	68 (21.3)
County level or other	48 (15.1)
Living status, n (%)	
Living alone	53 (16.6)
Living with family	212 (66.5)
Living with roommates	47 (14.7)
Other	7 (2.2)
Tried to quit smoking before, n (%)	
Yes	187 (58.6)
No	132 (41.4)
Used other health-related app, n (%)	
Yes	68 (21.3)
No	251 (88.7)
Downloaded smoking cessation app before, n (%)	
Yes	14 (4.4)
No	305 (95.6)
Average cigarettes per day, mean (SD)	15 (12)

[a]A provincial capital is the city exercising primary status in a state, province, or an autonomous region, usually as its seat of local government.

[b]A prefecture-level city is an administrative division of China, ranking below a province and above a county in China's administrative structure.

Table 2. Demographic characteristics of health care providers (N=224).

Variable and category	Statistics
Gender, n (%)	
Male	40 (17.9)
Female	184 (82.1)
Age (years), n (%)	
≤25	31 (13.8)
26-35	104 (46.4)
≥36	89 (39.7)
Occupation, n (%)	
Physician	81 (36.2)
Hotline worker	1 (0.4)
Health education commissioner	13 (5.8)
Social worker	16 (7.1)
Nurse	101 (45.1)
Others	12 (5.4)
Education level, n (%)	
High school or below	3 (1.3)
Bachelor’s or senior college degree	161 (71.9)
Master’s degree or above	60 (26.8)
Nationality, n (%)	
Han	218 (97.3)
Other	6 (2.7)
Living region, n (%)	
Metropolis (ie, Beijing, Shanghai, and Guangzhou)	123 (54.9)
Provincial capital	66 (29.5)
Prefecture-level city	35 (15.6)

Table 3. Providers' attitudes and beliefs about smoking cessation apps.

Items	Completely agree, n (%)	Somewhat agree, n (%)	Neutral, n (%)	Somewhat disagree, n (%)	Completely disagree, n (%)
Many of my clients or patients use mHealth[a] to manage their health.	67 (29.9)	82 (36.6)	65 (29.0)	7 (3.1)	3 (1.3)
mHealth apps hold promise as a tool to help people stop smoking.	79 (35.3)	91 (40.6)	46 (20.5)	8 (3.6)	0 (0.0)
There is good empirical evidence that stop smoking apps can help people quit.	53 (23.7)	100 (44.6)	59 (26.3)	11 (4.9)	1 (0.4)
As a clinician, I would recommend a stop smoking app to my patients or clients trying to quit.	85 (37.9)	74 (33.0)	56 (25.0)	6 (2.7)	3 (1.3)
Effective stop smoking apps are widely available for smokers.	55 (24.5)	57 (25.4)	91 (40.6)	19 (8.0)	2 (0.9)
If there were an app that allowed me to track my client or patients' progress quitting smoking, I would use it as a clinician.	107 (47.7)	78 (34.8)	39 (17.4)	0 (0.0)	0 (0.0)
If there were an empirically validated stop smoking app, I would recommend it.	144 (64.3)	60 (26.8)	20 (0.1)	0 (0.0)	0 (0.0)

[a]mHealth: mobile health.

Results

Demographic Characteristics of Samples

A total of 319 current smokers (average cigarettes per day: mean 15, SD 12) and 224 health care providers were included in the analysis. Table 1 shows the descriptive statistics for the demographic characteristics of smokers, 92.5% (295/319) individuals were male with an average age of 34.8 years and 7.5% (24/319) were female with an average age of 23.5 years. Almost half of the smokers (148/319, 46.4%) were employed full time, whereas 12.5% (40/319) of them were students. The education level of 85.6% (273/319) of smokers was bachelor's degree and above. The fraction of smokers from provincial capitals and higher-level cities (which includes metropolis, provincial capital, prefecture-level city, and county level [22,23]) was 63.6% (203/319). The fraction of smokers living with others was 81.2% (259/319). Only 4.4% (14/319) of the smokers had ever downloaded a smoking cessation app.

Regarding the health care providers (Table 2), 82.1% (184/224) were female, 39.7% (89/224) were aged more than 35 years, 45.1% (101/224) were nurses, and 36.2% (81/224) were physicians; the education level of 98.7% (221/224) of health care providers was bachelor's degree or above.

Providers and Smokers' Attitudes and Ratings About Smoking Cessation App and Features

Providers' attitudes and beliefs about mHealth cessation apps are summarized in Table 3. Most health care providers agreed (somewhat or completely) that mHealth apps hold promise for helping people quit smoking (170/224, 75.9%) and would recommend them to their clients (159/224, 70.9%), especially if the program was empirically validated (204/224, 91.1%). Half of the health care providers thought that effective cessation apps currently exist (112/224, 49.9%).

Smokers' and health care providers' ratings of important smoking cessation app features are compared in Table 4 and Figure 1. As can been seen in Table 4, more than 60% of smokers rated the following 4 features as very or extremely important: allow sharing the process of smoking cessation with family members and friends (216/319, 67.7%), help smokers track their progress (such as amount of smoking per day; 213/319, 66.8%), help with the side effects of medications and nicotine withdrawal symptoms (201/319, 63.0%), and adapt to ongoing needs and interests of smokers (194/319, 60.8%). Almost half of the smokers (136/319, 42.6%) rated the features of *allow smokers to disclose information on microblogs, WeChat, or other social networking sites* and *include games or entertainment projects* as not at all important. In contrast, the majority of health care providers rated all the features as very or extremely important except: includes games or entertainment (91/224, 40.6%) and includes videos about quitting smoking (52/224, 23.2%).

As shown in Figure 1, health care providers rated nearly all the features as more important than the smokers did. The only 3 exceptions were the apps being lower cost or free, being research tested, and allowing sharing the process of smoking cessation with family members and friends. Among the 5 domains, smokers rated *reputation* as more important than *communication*, *privacy and security*, and *content and user experience*. Moreover, *content and user experience* and *cost* were both significantly more important than *privacy and security* to smokers. In contrast, health care providers rated the domain of *privacy and security* more important than those of *communication* and *content and user experience*. Moreover, they rated the feature of *confidentiality of information* as the most important feature for cessation apps.

Among the subgroups of smokers defined by the demographic characteristics, the reader can see in the Multimedia Appendix 1 the statistically significant differences in the 5 overall domain

ratings between smokers younger than 35 years and those older than 36 years. There were no overall differences in the ratings by these subgroups we examined: occupation, education, nationality, living region, living status, tried to quit smoking in the past, used another health-related app in the past, and downloaded a smoking cessation app in the past.

Figure 1. Comparison of providers' and smokers' ratings of important features.

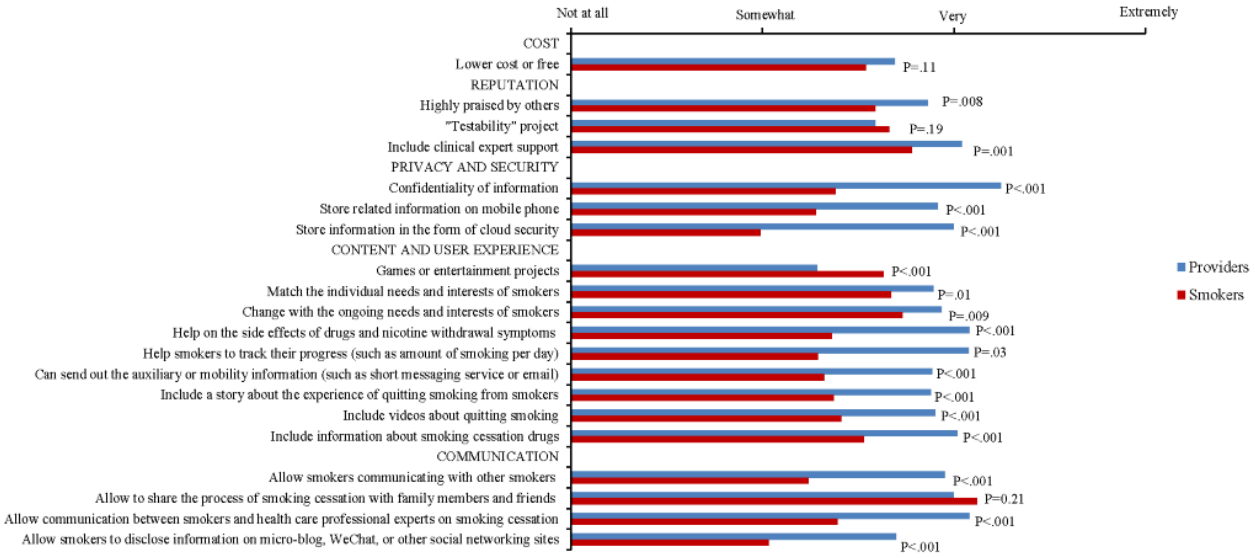


Table 4. Smokers and health care providers' ratings about smoking cessation apps features.

Features	Smokers, n (%)			Health care providers, n (%)		
	Not at all	Somewhat	Very or extremely	Not at all	Somewhat	Very or extremely
Allow sharing the process of smoking cessation with family members and friends	33 (10.3)	70 (21.9)	216 (67.7)	5 (2.2)	48 (21.4)	171 (76.3)
Help smokers track their progress (such as amount of smoking per day)	54 (16.9)	52 (16.3)	213 (66.8)	6 (2.7)	40 (17.9)	178 (79.5)
Help with the side effects of medications and nicotine withdrawal symptoms	58 (18.2)	60 (18.8)	201 (63.0)	7 (3.1)	42 (18.8)	175 (78.1)
Adapt to ongoing needs and interests of smokers	59 (18.5)	66 (20.7)	194 (60.8)	13 (5.8)	52 (23.2)	159 (71.0)
Keeps information private	70 (21.9)	61 (19.1)	188 (58.9)	8 (3.6)	29 (12.9)	187 (83.4)
Include clinical expert support	74 (23.2)	60 (18.8)	185 (58.0)	8 (3.6)	45 (20.1)	171 (76.4)
Match individual needs and interests of smokers	67 (21.0)	70 (21.9)	182 (57.1)	9 (4.0)	54 (24.1)	162 (71.8)
Highly rated by others	69 (21.6)	79 (24.8)	171 (53.6)	7 (3.1)	69 (30.8)	149 (66.1)
Allow communication between smokers and health care professional expert on smoking cessation	68 (21.3)	85 (26.7)	166 (52.0)	4 (1.8)	40 (17.9)	180 (80.4)
Low cost or free	72 (22.6)	90 (28.2)	157 (49.2)	15 (6.7)	81 (36.2)	128 (57.2)
Allow smokers to communicate with other smokers	83 (26.0)	89 (27.9)	147 (46.1)	6 (2.7)	53 (23.7)	165 (73.7)
Can send out auxiliary or mobility information (such as short message service or email)	76 (23.8)	99 (31.0)	144 (45.1)	8 (3.6)	64 (28.6)	152 (67.9)
Include stories about the experiences of quitting smoking from smokers	87 (27.3)	93 (29.2)	139 (43.6)	8 (3.6)	61 (27.2)	155 (69.2)
Include information about smoking cessation medications	90 (28.2)	91 (28.5)	138 (43.3)	7 (3.1)	46 (20.5)	171 (76.3)
Stores information on phone	77 (24.1)	107 (33.5)	135 (42.3)	12 (5.4)	57 (25.4)	155 (69.2)
Stores information in secure cloud	96 (30.1)	98 (30.7)	125 (39.2)	12 (5.4)	50 (22.3)	162 (72.3)
Program is *research tested*	96 (30.1)	104 (32.6)	119 (37.3)	16 (7.1)	91 (40.6)	117 (52.2)
Include videos about quitting smoking	108 (33.9)	99 (31.0)	112 (35.1)	7 (3.1)	60 (26.8)	157 (23.2)
Allow smokers to disclose information on microblogs, WeChat, or other social networking sites	136 (42.6)	87 (27.3)	96 (30.1)	24 (10.7)	68 (30.4)	132 (58.9)
Include games or entertainment	136 (42.6)	88 (27.6)	95 (29.8)	58 (25.9)	75 (33.5)	91 (40.6)

Discussion

Principal Findings

As far as we know, this was the first study to assess Chinese providers' and smokers' preferences for features of smoking cessation apps. The results offer researchers and developers guidance about what features might increase engagement with future smoking cessation apps [8]. Smokers rated *reputation* as more important than *communication*, *privacy and security*, and *content and user experience*. In traditional Chinese culture, the choice of products or services comes largely from the recommendation of friends or experts [23,24]. They are often unwilling to try new services without a reference from a trusted source [25]. Therefore, source expertise (eg, includes clinical expert support) and source trustworthiness (eg, highly praised by others and *research tested*) were rated very high among features of smoking cessation apps. Moreover, *content and user experience* and *cost* were both significantly more important than *privacy and security* to smokers.

In contrast, providers rated the domain of *privacy and security* more important than those of *communication* and *content and user experience*. Moreover, they rated the feature of *confidentiality of information* as the most important feature for cessation apps, which is consistent with the opinions of American providers [16]. The importance of privacy and security likely reflects the obligation of providers to protect the privacy of patients. Furthermore, providers rated the ability of an app to help smokers track their progress as well as obtain information about smoking cessation medications, side effects of nicotine withdrawal symptoms, and support and communication with clinical expertise as very important. As most smoking cessation apps in the Chinese market do not

currently provide these functionalities [10], it would seem important to add them in future apps. In contrast with American providers [16], Chinese providers rated games or entertainment projects as the least important.

In the content and user experience domain, having the app change to meet ongoing needs and interests was rated highly both by smokers and providers. This suggests that future smoking cessation apps should design content to include adaptively tailored features (eg, weekly graph based on the personal data smokers enter). The value of adaptive content is supported by a variety of behavior change theories [26,27].

Providers were generally supportive of apps for smoking cessation. The majority stated that they would recommend them and they would use them to track their patients' progress. Interestingly, a substantial fraction (49.9%) of health care providers thought that effective cessation apps currently exist, a belief that is not supported by the nascent research on these apps [28]. This may mean that this misperception is leading providers to incorrectly recommend apps for which there is no empirical support. Future app designs should be very clear about the current evidence to date on their effectiveness.

Comparing the results of this study of Chinese smokers with the previous study of US smokers reveals some intriguing contrasts. Specifically, security was rated as the least important domain by Chinese smokers in our study, whereas smokers in the US study rated that as the most important feature [16]. This contrast may be because smoking is not a private behavior in China: it functions as a way to make social connections, make smooth social interactions, and express one's social and economic position [29]. In contrast, smoking in America is highly stigmatized, which can have iatrogenic effects on quitting smoking [30]. Moreover, studies showed that Chinese people are more open and less sensitive about privacy [31]. According to a report published by market research firms (Experian and International Data Corporation) [32], people's attitudes toward Web-based privacy vary widely from country to country, with Chinese people being among the most willing to sacrifice privacy for safety and convenience. Therefore, privacy in quitting smoking would be highly valued for American smokers but not for Chinese smokers. Moreover, a second contrast with the US study was that Chinese smokers and providers of this study rated *allowing sharing of the process of smoking cessation with family members and friends* (communication domain) as the most important feature (P>.05), whereas the American smokers and providers rated that as relatively unimportant. This contrast likely reflects differences in Chinese values of family and close relationships, with behavior change involving the collective input of those people. Loved ones are seen as key sources of support to change a health habit in China [33].

Design Recommendations

The results yield a number of recommendations for the design of a smoking cessation app for Chinese smokers. Features rated very highly by both the smoker and provider samples suggest that an app should focus on (1) facilitating support for cessation from family and friends, (2) tracking their progress with quitting smoking, (3) getting help with the side effects of medications and nicotine withdrawal, and (4) being tailored to their unique needs. Results from the provider sample also suggest that an app should focus on protecting the privacy and confidentiality of the user as such a feature may make it more likely for providers to recommend the app to their patients. At the same time, these features rated consistently lower by both the smoker and provider samples suggest that an app for Chinese smokers should not (1) include games or entertainment and (2) allow disclosure on microblogs or WeChat.

Strengths and Limitations

This study has some notable strengths. This was the first study to investigate the opinions of Chinese health care providers and smokers on features in a quit smoking smartphone app. This study has limitations that can be addressed in future studies. First, as there is a broad diversity of smokers and providers in China, it is not possible to know to what extent the current sample's results can be generalized to the broader population of Chinese smokers or providers. Second, what smokers say they prefer may be different from what they actually will find useful once an app is prototyped. Thus, this study needs to be viewed as an important starting point in the design process. In-depth qualitative interviews and presenting early design sketches to smokers and providers would be the natural next steps in the process of creating a specific app.

Conclusions

Chinese health care providers expressed strong support of smoking cessation apps and indicated that nearly all features were important. Contrary to a similar study of US smokers and health care providers, Chinese smokers and providers highly value a smoking cessation app's reputation and ability to communicate with family members and friends as important features, whereas Chinese smokers rated privacy and security as less important. Therefore, the design of future smoking cessation and health behavior change apps should consider cultural differences and the perspectives of both health care providers and smokers.

Acknowledgments

This work was supported by a grant from the Tsinghua University Education Foundation (Global Health Program, project number 202400025). JB's time on this study was supported by the National Cancer Institute Grant R01 CA 192849-01A1.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ratings of features among subgroups with different ages (%).
[PDF File (Adobe PDF File)68 KB - mhealth_v7i10e12200_app1.pdf]

References

1. Pierce JP, Messer K, White MM, Kealey S, Cowling DW. Forty years of faster decline in cigarette smoking in California explains current lower lung cancer rates. Cancer Epidemiol Biomarkers Prev 2010 Nov;19(11):2801-2810 [FREE Full text] [doi: 10.1158/1055-9965.EPI-10-0563] [Medline: 20852009]
2. Hickey P, Evans DS. Smoking in Ireland 2014: Synopsis of Key Patterns. Ireland: Health Service Executive Publishing; 2015.
3. Germain D, Durkin S, Scollo M, Wakefield M. The long-term decline of adult tobacco use in Victoria: changes in smoking initiation and quitting over a quarter of a century of tobacco control. Aust N Z J Public Health 2012 Feb;36(1):17-23. [doi: 10.1111/j.1753-6405.2011.00805.x] [Medline: 22313701]
4. Nan Y, Xi Z, Yang Y, Wang LL, Tu MW, Wang JJ, et al. [The 2015 China adult tobacco survey: exposure to second-hand smoke among adults aged 15 and above and their support to policy on banning smoking in public places]. Zhonghua Liu Xing Bing Xue Za Zhi 2016 Jun;37(6):810-815. [doi: 10.3760/cma.j.issn.0254-6450.2016.06.014] [Medline: 27346107]
5. Zhu SB, Meng RG. [The implementation of MPOWER control strategy, to reduce the burden of cardiovascular disease]. Chin J of Evid Based Cardiovasc Med 2017;9:892-894. [doi: 10.3969/j.issn.1674-4055.2017.07.42]
6. Yang Y, Nan Y, Tu MW, Wang JJ, Wang LL, Jiang H. [Summary of China adult tobacco survey report 2015]. Chin J Health Manag 2016 Apr 1;10(2):85-87. [doi: 10.3760/cma.j.issn.1674-0815.2016.02.002]
7. National Health and Family Planning Commission of China. China clinical smoking cessation guideline. Chin J Health Manag 2016 Jul 4;10(2):2016 [FREE Full text] [doi: 10.3760/cma.j.issn.1674-0815.2016.02.003]
8. Fingrut W, Stewart L, Cheung KW. Choice of smoking cessation counselling via phone, text, or email in emergency department patients. Prev Med Rep 2016 Dec;4:597-600 [FREE Full text] [doi: 10.1016/j.pmedr.2016.10.010] [Medline: 27843760]
9. Statistical Reports - CNNIC. Statistical Report on the Development of the Thirty-Seventh China Internet Network URL: http://cnnic.cn/gywm/xwzx/rdxw/2015/201601/W020160122639198410766.pdf [accessed 2018-05-01]
10. Cheng F, Xu J, Su C, Fu X, Bricker J. Content analysis of smartphone apps for smoking cessation in China: empirical study. JMIR Mhealth Uhealth 2017 Jul 11;5(7):e93 [FREE Full text] [doi: 10.2196/mhealth.7462] [Medline: 28698170]
11. Consumer Health Information Corporation. 2008. Motivating Patients to Use Smartphone Health Apps URL: http://www.consumer-health [accessed 2018-05-06]
12. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
13. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GW, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
14. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
15. Ubhi HK, Kotz D, Michie S, van Schayck OC, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171 [FREE Full text] [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]
16. McClure JB, Hartzler AL, Catz SL. Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. JMIR Mhealth Uhealth 2016 Feb 12;4(1):e17 [FREE Full text] [doi: 10.2196/mhealth.5181] [Medline: 26872940]
17. Paay J, Kjeldskov J, Skov M, Rasmussen S. Understanding Individual Differences for Tailored Smoking Cessation Apps. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI'15; April 18-23, 2015; Seoul, Republic of Korea p. 1699-1708. [doi: 10.1145/2702123.2702321]
18. Ploderer B, Smith W, Howard S, Pearce J, Borland R. Introducing the Ambivalent Socialiser. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 5-10, 2012; Austin, Texas, USA p. 1511-1514. [doi: 10.1145/2207676.2208613]
19. Technology - Tencent. Tencent's Total Revenue for the First Quarter of 2016 Was 31.995 Billion Yuan, a Year-on-Year Increase of 43% URL: http://tech.qq.com/a/20160518/067853.htm [accessed 2018-06-01]
20. Beijing Channel--People's Daily Online People's Daily. QQ Simultaneous Online Account Number Exceeded 200 Million Mobile Terminal Login Users Over 70% URL: http://bj.people.com.cn/n/2014/0414/c339741-20987026.html [accessed 2018-06-01]
21. Ren YZ. Shantou University. 2011. Sleep Characteristics of Children With Attention Deficit Hyperactivity Disorder URL: https://max.book118.com/html/2014/0531/8539100.shtm [accessed 2019-08-23]

22. Ma G, Kong LZ, Luan DC, Dechun L, Yanping H, Xiaoqi W, et al. The Descriptive Analysis of the Smoking Pattern of People in China. Chin J Prev Control Chronic Non-Commun Dis 2005;13(5):195-199. [doi: 10.3969/j.issn.1004-6194.2005.05.002]
23. Beijing Real Estate_Beijing Real Estate Network_Beijing Real Estate Information Network-Beijing Phoenix Network Real Estate. What Level is Your Hometown? Beijing Phoenix Network Real Estate URL: https://house.ifeng.com/pic/2018_05_17-38896385_0.shtml [accessed 2018-06-07]
24. Shi Y. An empirical study on the timing of online product recommendation and the selection of information sources for recommendation. In: University of Science and Technology of China. Hefei: University of Science and Technology of China; Aug 01, 2013:2013.
25. Chen PY, Wu SY, Yoon J. The Impact of Online Recommendations and Consumer Feedback on Sales. In: Proceedings of the International Conference on Information Systems. 2004 Presented at: ICIS'04; December 12-15, 2004; Washington, DC p. 12-15.
26. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
27. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015 Jan;41(1):68-73 [FREE Full text] [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
28. 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]
29. Pan ZF. Socioeconomic predictors of smoking and smoking frequency in urban China: evidence of smoking as a social function. Health Promot Int 2004 Sep;19(3):309-315. [doi: 10.1093/heapro/dah304] [Medline: 15306615]
30. Kim J, Cao X, Meczkowski E. Does stigmatization motivate people to quit smoking? Examining the effect of stigmatizing anti-smoking campaigns on cessation intention. Health Commun 2018 Jun;33(6):681-689. [doi: 10.1080/10410236.2017.1299275] [Medline: 28398092]
31. Hou CH, Pan JH. Privacy and intercultural communication-causes behind different privacy of the Chinese and the american. Yingyu Guangchang 2012 Mar 19;2012(1):117-118 [FREE Full text]
32. Sina Finance. Li Yanhong Talks About Big Data: Chinese People Are Not Sensitive, Willing to Use Privacy for Convenience URL: http://finance.sina.com.cn/meeting/2018-03-26/doc-ifysqfnf8646820.shtml [accessed 2018-03-16]
33. Ceng YZ, Li RM, Liu SZ. The influence of short message intervention on the success rate of smoking cessation after percutaneous coronary intervention. Qilu Nursing Journal 2014 May 28(9):121-122. [doi: 10.3969/j.issn.1006-7256.2014.09.073]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 13.09.18; peer-reviewed by G Wu, YA Hong, A Nomura, J Vilaplana, L Grepo; comments to author 23.03.19; revised version received 06.06.19; accepted 19.07.19; published 04.10.19.

Please cite as:
Xu J, Bricker J, Fu X, Su C, Wang P, Qi T, Cheng F
Design and Development of Smoking Cessation Apps Based on Smokers' and Providers' Perspectives in China: Survey Study
JMIR Mhealth Uhealth 2019;7(10):e12200
URL: https://mhealth.jmir.org/2019/10/e12200
doi:10.2196/12200
PMID:31588914

Original Paper

End User–Informed Mobile Health Intervention Development for Adolescent Cannabis Use Disorder: Qualitative Study

Kara Bagot[1*], MD; Elizabeth Hodgdon[1*], BA; Natasha Sidhu[1*], BA; Kevin Patrick[2*], MD; Mikaela Kelly[3*], BA; Yang Lu[4*]; Eraka Bath[3*], MD

[1]Department of Psychiatry, University of California, San Diego, La Jolla, CA, United States

[2]Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA, United States

[3]Neuropsychiatric Institute, University of California, Los Angeles, Los Angeles, CA, United States

[4]University of California, San Diego, La Jolla, CA, United States

[*]all authors contributed equally

Corresponding Author:
Kara Bagot, MD
University of California, San Diego
Department of Psychiatry
9500 Gilman Drive
MC 0405
La Jolla, CA, 92093
United States
Phone: 1 8582462553
Email: kbagot@ucsd.edu

Abstract

Background: The rates of cannabis use continue to increase among adolescents and the current interventions have modest effects and high rates of relapse following treatment. There is increasing evidence for the efficacy of mobile technology–based interventions for adults with substance use disorders, but there is limited study of this technology in adolescents who use cannabis.

Objective: The goal of our study was to elucidate elements of an app-based adjunctive intervention for cannabis cessation that resonate with adolescents who use cannabis.

Methods: Adolescents, aged between 14 and 17 years, who used cannabis were recruited from San Diego County high schools. Semistructured focus groups (6 total; N=37) were conducted to examine the ways in which participants used smartphones, including the use of any health behavior change apps, as well as to elicit opinions about elements that would promote engagement with an app-based intervention for adolescent cannabis cessation. An iterative coding structure was used with first cycle structural coding, followed by pattern coding.

Results: Themes that emerged from the analysis included (1) youth valued rewards to incentivize the progressive reduction of cannabis use, which included both nontangible rewards that mimic those obtained on social media platforms and prosocial activity-related rewards, (2) having the ability to self-monitor progression, (3) peer social support, (4) privacy and confidentiality discrete logo and name and usernames within the app, and (5) individualizing frequency and content of notifications and reminders.

Conclusions: Integrating content, language, interfaces, delivery systems, and rewards with which adolescents who use cannabis are familiar, engage with on a day-to-day basis, and identify as relevant, may increase treatment engagement and retention for adolescents in substance use treatment. We may increase treatment effectiveness by adapting and individualizing current evidence-based interventions, so that they target the needs of adolescents and are more easily incorporated into their everyday routines.

(JMIR Mhealth Uhealth 2019;7(10):e13691) doi:10.2196/13691

KEYWORDS

adolescent; cannabis; mobile health; treatment; smartphone

Introduction

Cannabis is the most prevalent drug of abuse among adolescents, with nearly 50% of 12th graders, 30% of 10th graders, and 15% of 8th graders reporting lifetime use [1]. Although the rates of use of many substances have been on the decline among youth, cannabis use has failed to decline [1]. Greater than 25% of adolescents who use cannabis meet criteria for cannabis use disorder (CUD) [2]. The likelihood of developing cannabis dependence is linearly associated with the frequency of use and inversely associated with age, such that early onset users are the most susceptible for later dependence [3]. This is of concern as adolescents' brains are developing and more susceptible to perturbations because of substance exposure. Early exposure to high levels of delta-9-tetrahydrocannabinol through cannabis use triggers repeated activation of the endogenous mesolimbic dopaminergic system, which may in turn lead to sensitization of this system and progressive enhancement of acquired susceptibility to psychiatric illness [4]. Early regular cannabis use is associated with psychosocial consequences that increase burden of illness and decrease functional outcomes [5,6], including increased likelihood of other illicit drug use and poorer academic outcomes [7].

To date, the evidence for efficacious interventions targeting problematic cannabis use, especially in youth, are sparse, with some behavioral interventions demonstrating short-term abstinence during active treatment, but high rates of relapse at follow-up. Specifically, randomized controlled trials of behavioral interventions for adolescent cannabis use show moderate effect sizes posttreatment for motivational enhancement therapy (MET), cognitive behavioral therapy (CBT), family support therapy, case management, and contingency management (CM), with few adolescents maintaining abstinence through follow-up [6,8]. In a study evaluating CM plus MET and/or CBT on cannabis use outcomes in adolescents, those youth assigned to active treatment demonstrated longer durations of continuous abstinence during treatment and higher rates of point prevalence posttreatment abstinence as compared with behavioral treatment alone [9]. Further study of adjunctive CM has yielded conflicting results [10]. In addition, 5-week and 12-week MET and/or CBT, family support with home visits, psychoeducation, case management and referral to self-help groups, adolescent community reinforcement approach (operational skills training+social systems support), and multidimensional skills therapy have all demonstrated initial within-intervention efficacy, with reduction of use and increased abstinence. However, posttreatment, adolescents experienced frequent cycles of recovery and relapse; many were unable to maintain abstinence and the majority continued to report substance-related problems at follow-up [11]. This suggests that continued monitoring and intervention following active treatment may be important in maintaining gains during treatment.

Studies suggest that the effectiveness of interventions may be increased by developing selective, highly specific interventions that are presented on interactive platforms and include content addressing specific risk factors in at-risk youth [12]. However, few interventions have done so to date [13]. Mobile phones are a highly promising and palatable intervention platform for adolescents as access to mobile technology is ubiquitous; nearly 75% of adolescents own or have regular access to a smartphone (85% African American and 71% Hispanic and Caucasian) and 91% access the internet through a smartphone. Furthermore, 90% of adolescents with mobile phones use text messaging, sending and receiving an average of 30 messages daily [14]. Minority adolescents appear more likely to own smartphones and use apps, and those who are most likely to own a smartphone are also those in need of behavioral change interventions [15]. Therefore, there is enormous potential for leveraging mobile technologies to widely disseminate interventions for adolescent substance use. Furthermore, the ubiquity of mobile technology use among African American and Hispanic teenagers, provides the opportunity for increasing substance use treatment and public health intervention among historically underserved populations.

Although, adolescents spend a significant amount of time using technology, are expert users by virtue of exposure and use from childhood, and interactive platforms have been shown to be effective for treatment delivery in youth [12,14], few mobile health (mHealth) interventions (internet/mobile phone–based platforms) have been developed for use in adolescent substance users and none for adolescent cannabis use [16]. As such, we look to the adult data for guidance in examining the potential applicability of these types of interventions for at-risk adolescents. *CBT4CBT* (a computer-based CBT program supplementing therapist-delivered CBT has been shown to have greater efficacy (more days abstinent and fewer positive drug screens) than therapist-delivered CBT alone, for adults with SUDs up to 6 months posttreatment [17,18]. Web-based self-help interventions have also been shown to be efficacious in reducing alcohol use in adults [19-21]. Limited study of Web- and computer-based interventions in cannabis-using young adults and adults have demonstrated efficacy in correcting misperceptions about use and increasing knowledge but demonstrate mixed results for cessation outcomes [22-25]. Adults in a study of the *Therapeutic Education System* (self-guided, Web-based CBT modules+CM for completing modules and abstinence), exhibited higher rates of abstinence and greater treatment retention at study completion as compared with treatment-as-usual (TAU) [26]. A Web-based, self-guided intervention called *Reduce Your Use: How to Break the CannabisHabit*, which comprises modules based on CBT, MI, and behavioral self-management, demonstrated significantly greater rates of retention and reduction of cannabis use, frequency and quantity, during intervention as compared with controls [27]. The *Quit the Shit* intervention (52 weeks; Web-based questionnaires, weekly cannabis use diary with feedback, baseline and posttreatment therapist *chat*) also demonstrated a reduction in frequency and quantity of cannabis use in adults [28].

The goal of this study is to increase our understanding of adolescent cannabis use behaviors in the context of near universal availability of mobile, wireless, and wearable technology, with the overall aim of developing a user-informed mHealth intervention for high-risk cannabis-using adolescents engaged in outpatient substance use treatment. Specifically, we

aimed to (1) examine adolescent use of various mobile platforms (ie, short message service [SMS] text, apps), social media, and wearable devices to determine the use of mobile technologies as sources of health and substance use related information, and the effect on risk-taking behaviors and behavioral change, and (2) examine salience of components of existing behavioral interventions for youth.

Methods

Recruitment

We employed a stratified purposeful sampling strategy. A total of 37 adolescents aged between 14 and 18 years of age were recruited from high schools and substance use treatment centers in San Diego County, and San Diego Unified School District (SDUSD), which serves >130,000 students and is the second largest district in California. The racial/ethnic makeup of the students in the SDUSD is 46.5% Hispanic, 23.4% Caucasian, and 10.2% African American. Eligibility criteria for the focus groups included being 14 to 18 years of age, owning a smartphone and/or having internet access, speaking fluent English, and cannabis use one or more times in the past 30 days. If eligible, adolescents were invited to a focus group held either on their school campus, in the outpatient treatment center, or in the lab.

A total of 37 adolescents were enrolled in one of 6 focus groups, as that was the number needed to reach saturation (themes were repeated, no new information obtained) [29]. Groups were stratified on the basis of age (early adolescence: 14-16 years/8th-10th grades; late adolescence: 17-18 years/11th-12th grades) so that younger adolescents would not feel intimidated to speak in the presence of older teens. Age stratification was also done on the basis of data that marijuana use differs significantly by adolescent age, and the belief that differing rates of use and exposure to methods and strains of marijuana use might influence their experiences and ideas about a cannabis intervention. The groups included a maximum of 10 and a minimum of 5 adolescents per group so that the groups were small enough to give each participant the opportunity to speak and share ideas but large enough for some diversity. Adolescents were compensated with a US $25 gift card and travel to/from the focus group site was provided, if needed. Institutional Review Board approval and a Certificate of Confidentiality from the National Institutes of Health were obtained.

Design

Focus groups were conducted in English and followed a semistructured interview guide developed by the study principal investigator, a child and adolescent psychiatrist. Participants answered questions regarding characteristics of their smartphones, how they use their phones, what they use them for (including apps downloaded and native apps), time spent on their phones, use of related mobile technologies such as activity trackers, and use of mobile technology and Web-based platforms for accessing, monitoring, and/or changing health or substance related behaviors. Participants also worked as a group to design, and arrive at a consensus among group members, a mock app-based cannabis intervention for cannabis-using adolescents. They were prompted to discuss aspects of interest, including the following:

1. Content
   - Health-related content for in-app messages and/or push notifications
   - Access to psychoeducation and prior measured personal data (eg, cannabis use/treatment adherence)
   - Static (same intervention for all users) versus dynamic (individualized/tailored) messaging
   - Theory: learning (temporal feedback on individual level data to enhance engagement and behavior change) and social cognitive (increased self-efficacy with goal setting and defined expectations, positively reinforce behavior change)
2. Usage
   - Frequency of content delivery (exposure) via prompts, cue-related, global positioning system (GPS) location triggered
   - Format of content: text, video, pictures, and/or links to external sources
3. Duration: active (brief, extended) with or without continuing care (time limited vs indefinite access)
4. Contact
   - Clinicians via SMS text and Skype/video for confirmation of abstinence at routine intervals and in response to psychiatric or high-risk behavioral concerns
   - Peer supports via in-app social media
5. Accessibility
   - Navigational features
   - Design of user interface
   - Environments (feasibility of continuous monitoring/access to intervention across academic, social, and familial settings).

Measures

Participants were administered 4 self-report questionnaires: the Customary Drinking and Drug Use Record [30] to measure current and past substance use, an internet and a social media addiction scale (based on the Bergen Facebook Addiction Scale) [31], a smartphone attachment index (scale from 1 to 10 of the query *How attached are you to your smartphone?* with 10 being the highest), and a social media engagement questionnaire (queries related to time spent and actions taken on various platforms, including Instagram, Snapchat, YouTube, and Facebook). Table 1 presents participant characteristics, mean cannabis use, and phone attachment. Group discussions provided qualitative reports of how adolescents use mobile technology to communicate and track health behaviors, in addition to app development input. Audio recordings were transcribed verbatim and coded individually.

Table 1. Participant characteristics.

Demographic information	Values
Gender (n=37), n	
Female	22
Male	15
Age (years), mean (SD)	16.86 (0.82)
Race (n=37), n	
White/Caucasian	32
Black/African American	2
Not reported	3
Ethnicity (n=37), n	
Hispanic/Latino	29
Not reported	1
Smartphone and cannabis use, mean (SD)	
Phone attachment index (1-10)	6.97 (2.00)
Number of days cannabis used in past 30 days (n=24)	3.04 (5.20)

Statistical Analysis

Qualitative software, ATLAS.ti (Scientific Software Development GmbH) was used to aid in data management for thematic analyses of transcripts and coding [32]. We used a team coding approach with iterative coding. Coding was framed by the focus group guide developed following individual interviews of high-risk cannabis-using teens that helped to triangulate information obtained from the focus groups. Themes were selected a priori to correspond to the guide. These themes included health behavior change technology and components of platform design for mobile technology intervention development. A total oftwoindependent coders (KB and EH) determined the coding structures, coded the 6 focus group transcripts, and engaged in intensive discussion of the coded transcripts to refine the coding guide and arrive at consensus in instances of discrepancy. Code cooccurrence tables were analyzed for connections between codes and participant statements. We ensured rigor by checking coding through qualitative software (ATLAS.ti); staff were trained in qualitative methods and coders arrived at a consensus greater than 90% of the time.

Results

Health Behavior Change Mobile App Use

Group participants spoke about their use of various smartphone apps, including mHealth apps aimed at changing health behaviors such as sleep or nutrition (Table 2). Participants indicated they had used at least one mHealth app since owning a smartphone. Health behavior change app use was discontinued because of loss of interest or motivation, or dissatisfaction with the features. Wearable activity tracking devices were less popular, viewed as *uncool* and *obsessive* by some, and only useful when a person was dedicated to physical activity (ie, belonging to a sports team).

Table 2. Smartphone use: health behavior change technology.

Health behavior change devices and apps	Quotes
Use of wearable devices	"I have the Apple watch that I use for running too. And the Fitbit."; "I have an Apple watch but I hardly ever go running."; "[I use] a watch...to track my steps."
Use of activity tracking apps	"I used to use the health app that [iPhone] comes with...See how much I've walked and stuff."; "It's called Lifesum...it tracks your calories for the day...I've seen some other people using and I was like, 'that's pretty cool.'"
Reason to use health apps or devices	
Positive feedback	"All of those big numbers."
Physical training	"Like [if] I'm going to train..."; "[Using a wearable device] depends if I'm doing a sport."
Interface	"I just liked the way it looked, the layout...when you go on there, they ask you a whole bunch of questions. And then they give you a certain amount of calories for your weight and your height and like your age. And then you put in what you eat throughout the day...And then it's like a rainbow at the top. And it's like a sunset at the top and then water—I don't know. it's just a cool setup, I guess."
Reasons not to use health apps or devices	
Pressure to workout	"Because it's too much exercise...because I would want to get big numbers."
Uncool	"It's sort of just inherently uncool to me."
Fear of losing device	"I feel like I'd lose it."
Discontinuation	
Inconvenient/no new insight	"[Sleep app] I don't like to have my phone by my bed because it lights up and stuff. It had to be like pretty close to you for it to track...I mean like it was [useful]. It told me if I was sleeping good or not. But I'm also a really good sleeper so it always was like deep sleep and I'm like okay."
Not necessary	"I just didn't feel the need to have it. It was something to do."
Lost interest	"Well, at first, I would read them because it would just pop up, but then after like I would read them and they would just keep popping up and I wouldn't care and I would just swipe up. So, I was like oh I didn't even read them. So, I deleted it."
Perceptions of those who use cannabis	"Fitbit is like a preppy thing. It's the kids that would be willingly spend $100 for something a phone app can do."; "I'm just like oh my God, what a nerd. I just do not want to be associated with someone who is like, 'Oh, I've got to get my steps in today.'"; "It's a little obsessive to me."; "At least they're motivated."

Mobile App Design Features

During the group discussion, the following topics were considered: (1) content, (2) usage, (3) duration, (4) contact, and (5) accessibility. Themes from each area were derived through coding of focus group transcripts and are summarized in the following tables. The most prominent themes among participants were rewards, privacy, self-monitoring, peer social support, and notifications.

Rewards

Participants discussed creating a reward system for cannabis use cessation (Table 3). Participants proposed a contingency system of incremental rewards for sustained abstinence. Monetary rewards, such as prepaid debit cards, gift cards, coupons, or discounts, were suggested by participants across groups. In addition, participants proposed in-app rewards, such as points, emojis, pictures, exclusive app features, and games, as an alternative to monetary prizes.

Table 3. Preferences and design recommendations for an mHealth app for cannabis cessation: reward system.

Rewards	Quotes
Types of reward	
In-app rewards	"I don't think it has to be anything special. You know how like on Snapchat you get the random trophies...No one really cares about them but it's kind of like I check it. And I'm like oh what did I get? What have I done?"; "...funny pictures."; "You get cooler stickers."; "It's like every time you say you 'no' [to using marijuana, you] get a coin...like a game...that you can customize."
Gift card/prepaid debit card	"Yeah, a gift card to Starbucks. Something that you can district you."; "You can't go to a [drug dealer] with a credit card."; "[Gift cards for] Wal-Mart. Target. Regular shit."
Cash	"Money...Ten bucks or fifteen. Twenty."
Alternative activities	"I feel like with the app you guys should put in like events are happening around us. So it takes our mind off. You know, if we want to go somewhere else and do something it will take our mind off of wanting to smoke."
Coupons	"Maybe even tying in stuff from like Groupon where it gives you discounts to places."; "I like how they did like coupons or I don't know just something small."; "Or a coupon for like iTunes or something."; "...you can achieve a discount."
Streak	"You should incorporate like the streaks to your app...Streaks make everyone want to have to do, and on Snapchat they have a little hourglass if you're about to lose your streak. And once you see the hourglass, you're like, oh you better snap them. It becomes really important."; "[No marijuana use] then you just get a sticker or something or a streak to like whatever."; "Streaks make everyone so much more into it."
Reward frequency	
Contingent reward for staying *clean*	"Depending on how many days you stay clean you get certain rewards. And if you report that you did smoke then it will deduct points or deduct rewards."; "Send you a notification like hey, you've been clean for a week. This is what you get, this amount of points. And two weeks you can get like double the points."; "It probably goes up every month...Every month you stay clean."; "...if you say two weeks in the app that you're clean and you haven't smoked like you could get one...You should get rewarded."
Frequency needed to stay motivated	"Once a week."; "Every day."; "I feel like a month is good."; "Yeah, you guys could do one every week, a different one every week and you get a little bit every day."

Privacy

Several issues were raised about the protection of users' private information. Privacy considerations are presented in Table 4. Participants agreed that the app name and logo should be discrete to minimize the chance of parents or friends learning about their cannabis use. Password or passcode protection was another common recommendation.

Attitudes toward the use of location tracking in the app were divided. Supporters of location-based features liked the prospect of regional activity recommendations and a localized social support system. Dissenters argued that privacy was more important than local connection and they felt uncomfortable being monitored. A consensus was reached that the app should allow users to control location permissions.

Table 4. Preferences and design recommendations for an mHealth app for cannabis cessation: privacy features.

Privacy	Quotes
Connection to other social media accounts	"I definitely think it shouldn't post [to] Facebook, 'I haven't smoked in two months!'"; "And a lot of the times, when you join an app they're like find your friends on here. And they download your friends from your contacts. I don't think you should do that."; "You definitely shzould be able to connect through Twitter, Facebook."
Concern about global positioning system (GPS) tracking	"It's just weird to know that somebody could be tracking you wherever you are."; "Yeah, you could be at like at a club and be like where the fuck are you at?...You work for the FBI or what?"; "I wouldn't want the GPS for privacy reasons, you know. Some people are weird."
Anonymous user account	"Pick a user name...So, you wouldn't have to use your actual name."; "I think a user name would be the best. Just somehow it doesn't track back to actually you, never knowing actually who it is."; "Yeah, it has to totally be private because if a person sees you are on there and she's like why is she on there, oh, they'll tell people about what you're doing."; "I think user names, but there should be like one where if you want a person to know who you are, [you could]."
Passcode/fingerprint access	"Like when you open your phone, you know, like how there's a password, you can have a password for the app."; "You can use your fingerprint to open it."; "That would be kind of gnarly to put your fingerprint for something you want to stop smoking weed...More like a password. Something you remember like a four digit."
Discreet app name and logo	"You'd want it to be kind of discreet because a lot of kids have their parents check their apps they download."; "The logo on the app shouldn't be a marijuana leaf...it should be a discreet logo if your parents look at your phone, they don't just see, like—and the name shouldn't be something like, 'I'm quitting marijuana.'"

Self-Monitoring Cannabis Use

Participants suggested incorporating a feature for users to track their cannabis use. Various tools for collecting user information were proposed, including pop-up prompts and open-ended journal entries (Table 5). Personalized feedback in the form of text messages or graphic representations of cannabis use over time were presented as a way for users to view their progress.

A few group members voiced a need for abstinence verification. Solutions included in-person urine screening, remote drug tests, or providing an electronic testing device that directly attaches to the user's smartphone.

Table 5. Preferences and design recommendations for an mHealth app for cannabis cessation: self-monitoring cannabis use.

Self-monitoring	Quotes
Tracking marijuana use	"I think but the main goal is just to have progress, it's just to know like you're staying clean. So, I feel like it's recording if you've done it and...like whenever you're pressured and you said no, because those are skills to build."; "I think if you did smoke one day it should ask you questions like how you felt differently than when you did."; "Or even how often you wanted to smoke but didn't...When it used to be like oh I wanted to smoke three times, but now I only wanted to smoke once. Like either way it's progress."
Journaling	"...you need to know how to...when to say no, and when is the right time to do something."; "You have like a notepad, where you can write like this thing happened on this certain day and that's why I did it...and what I could learn from that and what could I change for the next time something that could happen."
Personalized feedback	"When you put the questions like how do you feel when you smoke? And they give you a response you could be like...or you could feel that way if you do this? Like I feel relaxed, like stressed out. And then you can be like 'get a soccer ball and you'll feel relaxed.'"; "Yeah, maybe put it in there so they could see like your ratio of in a month or like a year like how many times did you smoke? And if it's affecting your health or anything."
Verifying abstinence	"Send in some stuff. You've got to send in some, I don't know, some pee."; "Or like at the end have a meeting with them..."; "Why not that little detector like the alcohol detector? You get the app and then they send it to you."
Providing outside professional resources	"Maybe the app could provide professional help...Like really, really good resources like therapists that they offer or something or that they know of around where they're at."; "You know how some apps have the help, you know, like if they really want help. You put that under like oh here's some places that you can go to get help or something like that...A hotline."; "If they don't want to stop or something, that's when the hotline could come in. And then you like offer that talk. And then you talk to the person. And then if you see no hope then you tell them to talk to somebody else. That's when you offer the best help."

Peer Social Support

Every group proposed a social network type of messaging to anonymously connect with other app users who were actively trying to stop cannabis use. There was overall approval for the inclusion of social contact with other app users, with precautions taken to protect anonymity and location disclosure. Participants felt that including a social component served as a means of peer encouragement and distraction when cravings arise. Social support features are described in Table 6.

Notifications

Participants expressed a desire to have control over the frequency and content of the app notifications (Table 7). The upper limit mentioned by group participants was 3 times a day, though most preferred no more than once a day. When speaking about other mobile app use, participants expressed annoyance with receiving too frequent notifications and subsequently ignoring or turning off all notifications from the app, or deleting the app.

Table 6. Preferences and design recommendations for an mHealth app for cannabis cessation: peer social support.

Peer social support	Quotes
Talking to other app users	"Being able to talk to someone else maybe. But create a user name and you don't have to use your real name. That way if there's two people who are trying to quit, they can get together and like text each other."; "Part of the reason a lot of people don't stop is because they have a certain group of friends that are doing it. So, if they met someone else who was trying to stop, they would be like oh let's go do this instead of smoking with other friends. Create new friendships...Learn how to do other stuff with new people."; "There should be a forum for everyone to be able to join in on. And then separately it's like users, you can just click on them. And then if they have the option of being able to talk to other people because they want then okay, they should be able to."; "I think it would be really good to talk anonymously to teenagers, like, about trying to quit...Or you can put you in like a random chat..."; "And maybe you can join, like, a group? Like, a group so if you don't want to be with the others..."
Location-based chats	"I think it would actually be pretty cool if we have like location-wise, like, people around you, like in your city, it would even be pretty interesting to add to that."; "Let's just say San Diego has its own group chat and then there's, like, L.A. and there's, like, basically the big, major cities."; "There should be a thing like these people are in your location type of thing like when you have your location on...like these people who are also using the app are around you."

Table 7. Preferences and design recommendations for an mHealth app for cannabis cessation: notifications.

Notifications	Quotes
Notification delivery	
Frequency	"[...once a day would be the most frequent?] Or two times a day maybe."; "Maybe three. When you wake up, before you go to sleep and the middle of the day."; "I don't think it should be like constant reminder of like you've been clean for this amount because that will keep you thinking of like still weed. So, it should have it probably once every couple of weeks. And then probably have other motivational things."
Timing	"Like when you're out partying or something like that. You know, you're just in the moment of feeling good. Sometimes people smoke when they're in the vibe too...Or maybe get more notifications at that point on."
User control	"There should be a setting where you can like say you can press I want to check my own notifications. Or I want notifications every two weeks or like every day. Something like that."; "[You want control over the content of the notifications…] And how often you get them."
Notification content	
Reminders of progress	"It would be kind of cool if every three or so days or if it got like a long time like every five like how many days you're clean because it's kind of like a reminder. Like okay I'm clean I can't mess this up, I need to keep on this...it's three days. Let's get to five."; "There should be two different types of reminders and one of them is just to motivate you not to. And then the other one is just check in for tonight and it reminds you at night."
User control	"I would want independent control...cause I would know I could change it if I wanted to."; "They should also have the option of what kind of information the notification they're giving out. So, for example, some people probably don't...wouldn't want for the app to be telling them how bad it is to smoke or something. But some others would find it more motivational too, for the app to tell you oh, this is bad for you because of this or whatever."; "Maybe for inspirational...Maybe it could be designed into how you want it be. Because reminder quotes people will find irrelevant. So, when you're getting an inspirational quote you can kind of design it to how you want it to be."
Location-based notifications	"You'll get a notification like oh there's a concert happening in the park. Or like oh this museum is not charging today. Something like that. It all depends on where you're at. So it's around your surroundings."

Discussion

Principal Findings

We queried adolescents who use cannabis about specific aspects of a mobile technology–based intervention, including content, duration, user interface, accessibility, support (peer and clinician), and current/past use of health behavior change apps, to inform development of a cannabis cessation intervention that would increase treatment engagement by youth in substance use treatment.

Thetwomost common themes in intervention development that were expressed by participants were (1) monitoring reduction of cannabis use over time and (2) providing rewards for successful reduction and cessation. Adolescents indicated that integration of CM would be appealing. Previous research suggests that CM can improve effectiveness of substance use treatment for youth and increase engagement in treatment [33-36], including for youth with CUDs [37]. Surprisingly, participants in our study indicated that nonmonetary rewards would be the most rewarding; this included receiving emojis, similar to *streaks* in Snapchat, which many expressed would be motivating enough to maintain abstinence. Participants also indicated that being rewarded with *Groupons* for prosocial activities, such as movie theater or concert tickets, where one would be unable to use substances, would be appealing. Participants also expressed preference for receiving information and/or tickets for activities and events specific to their region, as well as rides to and from these activities and indicated that they would approve of integrating geolocation tracking services into the intervention to individualize rewards. Ironically, this is despite research showing that 46% of adolescents are likely to turn off or disable geolocation features associated with downloaded apps because of concerns of privacy [15]. The difference between our findings and those previously reported may be because of expectations of privacy starting at the point of download. For our purposes, there is an expectation that personal information would be collected to treat a disorder, whereas adolescents in the Pew Research study indicated downloading free apps for entertainment purposes where there is no expectation that personal information is needed to reach the desired outcome. Furthermore, adolescents may be more willing to share personal information, including location, when it results in a reward. Other studies have found that privacy and confidentiality were of less concern to high-risk adolescents if addressed at the beginning of a mHealth intervention [38]. Behavioral interventions for adolescent substance use treatment incorporate encouragement of prosocial activities to avoid substance use, with increased efficacy of intervention, if adolescents are able to increase engagement in these activities. The adolescents enrolled in this study also independently identified engagement in alternative, prosocial activities as a way to decrease use, and indicated updates to traditional thoughts about types of activities (eg, school-based sports or clubs) to technology-accessed activities specific to their lives and communities.

Privacy was an interesting theme that emerged in an isolated context, but also within discussion of other content areas. When asked directly about privacy and others' potential to have access to participants' data, knowledge of their involvement in the intervention, or location-based information available to researchers/clinicians for intervention delivery, participants almost uniformly had negative views. However, with regard to rewards and peer resources, participants voluntarily and spontaneously proposed offering personal information and inclusion of location-based services to gain psychosocial and other intangible rewards, as well as tangible rewards.

The theme of individualizing the app was also discussed in other domains. Individualization of mHealth interventions has been identified in previous studies as important to high-risk adolescents seeking treatment [38]. In our study, adolescents indicated their desire to individualize all aspects of the app, including the user interface, frequency, timing, tone and content of notifications and messages, and the level of clinician, peer, or text-based responsiveness to fluctuations in mood states and substance use. Dynamic interventions that address these factors are currently lacking from mHealth interventions, especially those tailored to youth. The majority of mHealth interventions for substance using and/or at-risk youth are static, text message–based interventions that do not address differences in baseline severity of substance use, psychiatric symptoms, and/or adapt to changes in these behaviors or symptoms over time.

Themes identified by adolescents who use cannabis, the target population for this app-based cannabis use intervention, are consistent with research suggesting that intervention efficacy may be increased by developing selective, highly specific interventions incorporating content that addresses risk factors in at-risk youth and are presented on interactive platforms [39].

Limitations

There are several limitations to this study. As we sampled adolescents who use cannabis from one city in Southern California, it was not representative of the broader adolescent population. In addition, we had a low number of racial and ethnic minorities in our sample, further limiting generalizability. With regard to uses of mobile technology, we did not query adolescents about alternate platforms for intervention other than an app delivered via smartphone. Although the literature shows that adolescents' primary mode of mobile technology use are smartphones, we did not query participants about use or acceptability of other types of technologies. Finally, we did not query youth on general methods of access to the app, as the goal is to integrate this into extant treatment paradigms for users engaged in substance use treatment. Despite these limitations, the qualitative approach allowed us to further probe the complexities and context of components of intervention development to understand adolescent technology use, and motivations for appeal of component content, more in-depth.

Future Directions

This app, incorporating mobile CBT, is being developed as an adjunct to TAU for adolescent cannabis users aged between 13 and 18 years who have 9 to 12 weeks of substance use treatment remaining at the time of study entry (to ensure we enroll youth before TAU impacts substance use). Participants will be recruited from local adolescent substance use treatment clinics.

The youth clinics provide case management and individual and group counseling intwotracks, outpatient drug-free groups (meet daily, Monday-Friday for 1.5 hours after school) and day care habilitative groups (intensive daily, Monday-Friday, half-day program), with adjunctive family therapy, parent and youth support meetings, and aftercare counseling, as needed. Random urine toxicology is done to confirm abstinence. Youth are enrolled in treatment for 3 months, with 1 month added to treatment length for each positive urine toxicology screen. An active cohort (TAU+biosensor) will be matched by primary substance of use to a cohort receiving TAU alone (matched comparison cohort data collected through electronic medical record [EMR]/chart review) in an external control design. We will assess and compare groups on the primary outcome of point prevalence abstinence (total number throughout treatment and number of consecutive weeks) for primary substance of use (substance for which the youth meets criteria for SUD), and secondary outcome of point prevalence abstinence for total number of substances used. These outcomes will be collected for the TAU group by reviewing their medical records. Collecting primary outcome comparison data through EMR/chart review will provide us with an important opportunity to identify systems-related considerations of cross-database information sharing, privacy, confidentiality, and other issues that must be resolved before launching a larger-scale study, in which obtaining collateral EMR/chart data will be essential to supporting treatment outcomes research trials.

Conclusions

Adolescent cannabis use initiation is linked to negative long-term health effects [38,39], with likely greater impact because of rising delta-9-tetrahydrocannabinol concentrations [40,41]. Technology-based substance use interventions appear promising in adult and college-aged populations, but few are adapted for youth, and none specifically targeted for adolescents with CUD. Behavioral interventions for adolescent cannabis use have demonstrated limited success, with high rates of relapse and loss of gains at follow-up [11,13]. mHealth interventions allow for examination of behaviors, ecological and mental states, environments, and social networks contributing to, and resulting from, adolescent substance use in real time, with near instantaneous response and intervention. Collectively, real-time data can shape targeted treatment around individual risk factors for problematic use. Furthermore, mHealth interventions have the potential to address current barriers to treatment, including cost, stigma, and access to providers with adequate training in utilizing/implementing behavioral interventions specifically targeting substance use, as well as increasing overall reach and scalability of interventions.

We have presented novel end user–informed data about the content, format, structure, privacy, and accessibility of an app-based substance use treatment for adolescents that may inform more successful interventions among this high-risk population. In this study, the sample of adolescents who use cannabis indicated a desire for an individualized app, with highly visual components consistent with apps that they already use and escalating rewards associated with individual progress. Furthermore, although response to use of geolocation services in the context of discussion of privacy was mixed, adolescents endorsed approval of sharing GPS information in the context of discussion pertaining to individualized rewards and connecting with peers. Identifying and developing intervention content on platforms highly utilized by target population, incorporating skill development (eg, saying no, coping with negative feeling, engagement in prosocial activities) via novel technological means, and integrating salient environmental rewards may increase intervention efficacy, and thus may improve substance use treatment outcomes. Further study of ethical and privacy implications of novel technological approaches are needed in this population. However, methods of ensuring that data are transmitted and stored in a manner that complies with Health Insurance Portability and Accountability (HIPAA) requirements are possible by ensuring that (1) all information from mobile to server will be via https/SSL (secure sockets layer), (2) stored data are encrypted and deleted when no longer required, (3) servers on the cloud are HIPAA compliant and regularly patched with security updates, stored in a secure facility, and (4) provision of a clear privacy policy and (5) strong passwords with expiration periods are mandated.

Acknowledgments

Susan Tapert, PhD; Suchitra Krishnan-Sarin, PhD; Dana Cavallo, PhD; Alexander Chang, MD; Emily Springfield, BA; 5K12 DA000357-17 (PIs: KB, EB); R25DA035163 (PIs: Carmen Masson, PhD, James Sorensen, PhD; subaward PI: KB).

This work was supported by the National Institutes on Drug Abuse—American Academy of Child and Adolescent Psychiatry Physician Scientist Program in Substance Abuse K12 Award (5K12 DA00035717) awarded to KB.

Conflicts of Interest

None declared.

References

1. Miech R, Johnston L, O'Malley P, Bachman J, Schulenberg J, Patrick M. Monitoring the Future. 2017. Monitoring the future national survey results on drug use: Volume I: Secondary school students URL: http://www.monitoringthefuture.org/pubs/monographs/mtf-vol1_1999.pdf

2. Wu LT, Brady KT, Mannelli P, Killeen TK, NIDA AAPI Workgroup. Cannabis use disorders are comparatively prevalent among nonwhite racial/ethnic groups and adolescents: a national study. J Psychiatr Res 2014 Mar;50:26-35 [FREE Full text] [doi: 10.1016/j.jpsychires.2013.11.010] [Medline: 24342767]
3. Gray KM. New developments in understanding and treating adolescent marijuana dependence. Adolesc Psychiatry (Hilversum) 2013;3(4):297-306. [doi: 10.2174/22106766030414012117321 5] [Medline: 25289370]
4. Laruelle M. The role of endogenous sensitization in the pathophysiology of schizophrenia: implications from recent brain imaging studies. Brain Res Brain Res Rev 2000;31(2-3):371-384. [Medline: 10719165]
5. Brewer TL, Collins M. A review of clinical manifestations in adolescent and young adults after use of synthetic cannabinoids. J Spec Pediatr Nurs Apr 2014;19(2):119-126. [doi: 10.1111/jspn.12057] [Medline: 24320158]
6. Copeland J, Swift W. Cannabis use disorderpidemiology and management. Int Rev Psychiatry Apr 2009;21(2):96-103. [doi: 10.1080/09540260902782745] [Medline: 19367503]
7. Hall WD. Cannabis use and the mental health of young people. Auzt N Z J Psychiatry 2006;40(2):105-113. [doi: 10.1080/j.1440-1614.2006.01756.x] [Medline: 16476127]
8. Patnode CD, O'Connor E, Rowland M, Burda BU, Perdue LA, Whitlock EP. Primary care behavioral interventions to prevent or reduce illicit drug use and nonmedical pharmaceutical use in children and adolescents: a systematic evidence review for the US Preventive services task force. Ann Intern Med 2014;160(9):612-620. [doi: 10.7326/M13-2064] [Medline: 24615613]
9. Stanger C, Budney AJ, Kamon JL, Thostensen J. A randomized trial of contingency management for adolescent marijuana abuse and dependence. Drug Alcohol Depend 2009;105(3):240-247. [doi: 10.1016/j.drugalcdep.2009.07.009] [Medline: 19717250]
10. Kaminer Y, Burleson JA, Burke R, Litt MD. The efficacy of contingency management for adolescent cannabis use disorder: a controlled study. Subst Abuse 2014;35(4):391-398. [doi: 10.1080/08897077.2014.933724] [Medline: 25010430]
11. Dennis M, Godley SH, Diamond G, Tims FM, Babor T, Donaldson J, et al. The cannabis youth treatment (CYT) study: main findings from two randomized trials. J Subst Abuse Treat 2004 Oct;27(3):197-213. [doi: 10.1016/j.jsat.2003.09.005] [Medline: 15501373]
12. Kumpfer KL. National Institute on Drug Abuse. 2015. Identification of Drug Abuse Prevention Programs: A Literature Review URL: http://archives.drugabuse.gov/about/organization/despr/hsr/da-pre/KumpferLitReview.html [accessed 2019-06-03] [WebCite Cache ID 78relsmZ3]
13. Champion KE, Newton NC, Barrett EL, Teesson M. A systematic review of school-based alcohol and other drug prevention programs facilitated by computers or the internet. Drug Alcohol Rev 2013 Mar;32(2):115-123. [doi: 10.1111/j.1465-3362.2012.00517.x] [Medline: 23039085]
14. Lenhart A. Pew Research Center. 2015. Teens, Social Media and Technology Overview 2015 URL: https://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2019-06-03] [WebCite Cache ID 78rf78i3F]
15. Madden M, Lenhart A, Gasser U. Pew Research Center. 2013. Teens and Mobile Apps Privacy URL: https://www.pewinternet.org/2013/08/22/teens-and-mobile-apps-privacy/
16. Kapitány-Fövény M, Vagdalt E, Ruttkay Z, Urbán R, Richman MJ, Demetrovics Z. Potential of an interactive drug prevention mobile phone app (once upon a high): questionnaire study among students. JMIR Serious Games 2018 Dec 4;6(4):e19 [FREE Full text] [doi: 10.2196/games.9944] [Medline: 30514697]
17. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Nuro KF, et al. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. Am J Psychiatry 2008 Jul;165(7):881-888 [FREE Full text] [doi: 10.1176/appi.ajp.2008.07111835] [Medline: 18450927]
18. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Rounsaville BJ. Enduring effects of a computer-assisted training program for cognitive behavioral therapy: a 6-month follow-up of CBT4CBT. Drug Alcohol Depend 2009 Feb 01;100(1-2):178-181 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.09.015] [Medline: 19041197]
19. Blankers M, Koeter MW, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341. [doi: 10.1037/a0023498] [Medline: 21534652]
20. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of e-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011 Jun 30;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]
21. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
22. Gulliver A, Farrer L, Chan JK, Tait RJ, Bennett K, Calear AL, et al. Technology-based interventions for tobacco and other drug use in university and college students: a systematic review and meta-analysis. Addict Sci Clin Pract 2015 Feb 24;10:5 [FREE Full text] [doi: 10.1186/s13722-015-0027-4] [Medline: 25928221]
23. Budney AJ, Fearer S, Walker DD, Stanger C, Thostenson J, Grabinski M, et al. An initial trial of a computerized behavioral intervention for cannabis use disorder. Drug Alcohol Depend 2011 May 1;115(1-2):74-79 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.10.014] [Medline: 21131143]

24. Elliott JC, Carey KB, Vanable PA. A preliminary evaluation of a web-based intervention for college marijuana use. Psychol Addict Behav 2014 Mar;28(1):288-293. [doi: 10.1037/a0034995] [Medline: 24731118]
25. Kay-Lambkin FJ, Baker AL, Lewin TJ, Carr VJ. Computer-based psychological treatment for comorbid depression and problematic alcohol and/or cannabis use: a randomized controlled trial of clinical efficacy. Addiction 2009 Mar;104(3):378-388. [doi: 10.1111/j.1360-0443.2008.02444.x] [Medline: 19207345]
26. Cochran G, Stitzer M, Campbell AN, Hu MC, Vandrey R, Nunes EV. Web-based treatment for substance use disorders: differential effects by primary substance. Addict Behav 2015 Jun;45:191-194 [FREE Full text] [doi: 10.1016/j.addbeh.2015.02.002] [Medline: 25697725]
27. Rooke S, Copeland J, Norberg M, Hine D, McCambridge J. Effectiveness of a self-guided web-based cannabis treatment program: randomized controlled trial. J Med Internet Res 2013 Feb 15;15(2):e26 [FREE Full text] [doi: 10.2196/jmir.2256] [Medline: 23470329]
28. Tossmann HP, Jonas B, Tensil MD, Lang P, Strüber E. A controlled trial of an internet-based intervention program for cannabis users. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):673-679. [doi: 10.1089/cyber.2010.0506] [Medline: 21651419]
29. 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]
30. Brown SA, Myers MG, Lippke L, Tapert SF, Stewart DG, Vik PW. Psychometric evaluation of the customary drinking and drug use record (CDDR): a measure of adolescent alcohol and drug involvement. J Stud Alcohol 1998 Jul;59(4):427-438. [Medline: 9647425]
31. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a Facebook addiction scale. Psychol Rep 2012 Apr;110(2):501-517. [doi: 10.2466/02.09.18.PR0.110.2.501-517] [Medline: 22662404]
32. Muhr T. Atlas.ti. 2018. URL: https://atlasti.com/
33. Branson CE, Barbuti AM, Clemmey P, Herman L, Bhutia P. A pilot study of low-cost contingency management to increase attendance in an adolescent substance abuse program. Am J Addict 2012;21(2):126-129. [doi: 10.1111/j.1521-0391.2011.00204.x] [Medline: 22332855]
34. Lott DC, Jencius S. Effectiveness of very low-cost contingency management in a community adolescent treatment program. Drug Alcohol Depend 2009 Jun 1;102(1-3):162-165. [doi: 10.1016/j.drugalcdep.2009.01.010] [Medline: 19250774]
35. Henggeler SW, McCart MR, Cunningham PB, Chapman JE. Enhancing the effectiveness of juvenile drug courts by integrating evidence-based practices. J Consult Clin Psychol 2012 Apr;80(2):264-275 [FREE Full text] [doi: 10.1037/a0027147] [Medline: W]
36. Godley MD, Godley SH, Dennis ML, Funk RR, Passetti LL, Petry NM. A randomized trial of assertive continuing care and contingency management for adolescents with substance use disorders. J Consult Clin Psychol 2014 Feb;82(1):40-51 [FREE Full text] [doi: 10.1037/a0035264] [Medline: 24294838]
37. Kamon J, Budney A, Stanger C. A contingency management intervention for adolescent marijuana abuse and conduct problems. J Am Acad Child Adolesc Psychiatry 2005 Jun;44(6):513-521. [doi: 10.1097/01.chi.0000159949.82759.64] [Medline: 15908833]
38. Ranney ML, Choo EK, Cunningham RM, Spirito A, Thorsen M, Mello MJ, et al. Acceptability, language, and structure of text message-based behavioral interventions for high-risk adolescent females: a qualitative study. J Adolesc Health 2014 Jul;55(1):33-40 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.12.017] [Medline: 24559973]
39. Kumpher K. National Institute of Drug Abuse. 2018. Identification of Drug Abuse Prevention Programs: A Literature Review URL: http://archives.drugabuse.gov/about/organization/despr/hsr/da-pre/KumpferLitReview.html#universal [accessed 2019-06-03] [WebCite Cache ID 78rfGLPTn]
40. ElSohly MA, Ross SA, Mehmedic Z, Arafat R, Yi B, Banahan 3rd BF. Potency trends of delta9-THC and other cannabinoids in confiscated marijuana from 1980-1997. J Forensic Sci 2000 Jan;45(1):24-30. [Medline: 10641915]
41. Freeman TP, Winstock AR. Examining the profile of high-potency cannabis and its association with severity of cannabis dependence. Psychol Med 2015 Nov;45(15):3181-3189 [FREE Full text] [doi: 10.1017/S0033291715001178] [Medline: 26213314]

Abbreviations

CBT: cognitive behavioral therapy
CM: contingency management
CUD: cannabis use disorder
EMR: electronic medical record
GPS: global positioning system
HIPAA: Health Insurance Portability and Accountability
MET: motivational enhancement therapy
mHealth: mobile health
SDUSD: San Diego Unified School District

SMS: short message service
TAU: treatment-as-usual

Edited by C Dias; submitted 13.02.19; peer-reviewed by M Harris, G Cutter, J Niu; comments to author 09.04.19; revised version received 14.05.19; accepted 19.05.19; published 04.10.19.

Please cite as:
Bagot K, Hodgdon E, Sidhu N, Patrick K, Kelly M, Lu Y, Bath E
End User–Informed Mobile Health Intervention Development for Adolescent Cannabis Use Disorder: Qualitative Study
JMIR Mhealth Uhealth 2019;7(10):e13691
URL: https://mhealth.jmir.org/2019/10/e13691
doi:10.2196/13691
PMID:31588909

Original Paper

Nurse-Driven mHealth Implementation Using the Technology Inpatient Program for Smokers (TIPS): Mixed Methods Study

Amanda C Blok[1,2], PhD, MSN, PHCNS-BC; Rajani S Sadasivam[3], PhD; Timothy P Hogan[3,4], PhD; Angela Patterson[3], MA; Nicole Day[5], MEd; Thomas K Houston[6], MPH, MD

[1]Veterans Affairs Center for Clinical Management Research, Veterans Affairs Ann Arbor Healthcare System, United States Department of Veterans Affairs, Ann Arbor, MI, United States

[2]Systems, Populations and Leadership Department, School of Nursing, University of Michigan, Ann Arbor, MI, United States

[3]Division of Health Informatics and Implementation Science, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[4]Veterans Affairs Center for Healthcare Organization and Implementation Research, Veterans Affairs Bedford Medical Center, United States Department of Veterans Affairs, Bedford, MA, United States

[5]University of Massachusetts Memorial Health Center, Worcester, MA, United States

[6]Learning Health Systems, Department of Medicine, Wake Forest University, Winston-Salem, NC, United States

Corresponding Author:
Amanda C Blok, PhD, MSN, PHCNS-BC
Veterans Affairs Center for Clinical Management Research
Veterans Affairs Ann Arbor Healthcare System
United States Department of Veterans Affairs
2215 Fuller Road, Mail Stop 152
Ann Arbor, MI,
United States
Phone: 1 734 845 3502
Email: amanda.blok@va.gov

Abstract

Background: Smoking is the leading cause of preventable death and disease, yet implementation of smoking cessation in inpatient settings is inconsistent. The Technology Inpatient Program for Smokers (TIPS) is an implementation program designed to reach smokers with a mobile health (mHealth) intervention using stakeholder-supported strategies.

Objective: The purpose of this study was to determine the impact of the TIPS implementation strategies on smoker-level engagement of the mHealth intervention during care transition.

Methods: We examined varying intensities (passive motivational posters only and posters + active nurse-led facilitation) of TIPS strategies on four hospital units located in two sites. Unit-level and smoker-level adoption was monitored during active implementation (30 weeks) and sustainability follow-up (30 weeks). Process measures reflecting the reach, effectiveness, adoption, implementation, maintenance (RE-AIM) framework, stakeholder reported adaptations of strategies, and formative evaluation data were collected and analyzed.

Results: For our smoker-level reach, 103 smokers signed up for the mHealth intervention in-hospital, with minimal decline during sustainability follow-up. While posters + nurse facilitation did not lead to higher reach than posters alone during active implementation (27 vs 30 signed up), it did lead to higher engagement of smokers (85.2% vs 73.3% completion of the full 2-week intervention). TIPS strategy adoption and fidelity varied by unit, including adoption of motivational posters (range: weeks 1 and 5), fidelity of posters (0.4% to 16.2% of posters missing per unit weekly) and internal facilitation of nurse training sessions (average of 2 vs 7.5 by site). Variable maintenance costs of the program totaled US $6.63 (US $683.28/103) per smoker reached. Reported family-member facilitation of mHealth sign-up was an observation of unintended behavior.

Conclusions: TIPS is a feasible and low-cost implementation program that successfully engages smokers in an mHealth intervention and sustains engagement after discharge. Further testing of nurse facilitation and expanding reach to patient family and friends as an implementation strategy is needed.

(JMIR Mhealth Uhealth 2019;7(10):e14331) doi:10.2196/14331

KEYWORDS

implementation strategy; telemedicine; mHealth, tobacco use cessation; care transition; patient transfer; smoking cessation; mobile health; smoking; tobacco

Introduction

Tobacco use is the primary preventable cause of death and disease in the United States [1]. Treating smoking-related illness costs $170 billion in direct medical costs and over $156 billion in lost productivity [1,2]. Furthermore, tobacco use is a major risk factor for many chronic illnesses that commonly result in hospitalization, such as cancer, heart and lung diseases, chronic obstructive pulmonary disease, and diabetes [1]. Patients who resume smoking upon discharge are more likely to be rehospitalized [3]. Hospitalization is a unique period of forced abstinence, and this is an opportunity to engage smokers and motivate them to be smoke free as they transition home after discharge [1]. However, continual engagement with tobacco cessation support can be a challenge when transitioning away from the clinical setting.

Mobile technology allows for on-hand support and is often willingly ported by users. More than 9 out of 10 smokers in the United States own a mobile phone, and the majority of patients in the hospital setting can access text-enabled phones [4,5]. The application of mobile health (mHealth) technology has been recognized as an evidence-based approach to tobacco cessation since 2011 [6], yet programs to connect smokers with this technology have not been implemented in hospitals.

While tobacco cessation interventions have been effective [7], they are often challenging for hospitals to implement and sustain [8]. While intensive interventions involving nurse-administered toolkits have increased abstinence [9], they are difficult to integrate into hospital staff workflow, leaving a gap in dissemination across hospital settings. Further, these nurse-administered interventions infrequently engage smokers after discharge. Hospital managers across the United States report a perceived lack of action addressing tobacco use in the hospital setting [10]. Hospital staff members face numerous barriers to addressing patient tobacco use, including time constraints, inadequate support, and limited training in tobacco cessation counseling [11,12]. Hospitals need implementation programs designed to address these barriers [8].

We developed an implementation program called the Technology Inpatient Program for Smokers (TIPS) to support the use of an mHealth intervention designed to continue engaging smokers during transition into the outpatient setting [13]. TIPS is a low-cost program that lets managers and staff employ intuitive, theory-driven strategies to reach people using a technology-supported behavioral intervention. The TIPS program consists of multiple strategies: (1) motivating mHealth intervention use through a promotional poster campaign and (2) activating nursing staff to facilitate patient sign-up. The purpose of this study was to determine the adoption and fidelity of TIPS strategies by units and measure smokers' engagement in the mHealth intervention.

Methods

Study Design

TIPS was evaluated at two sites using a phased implementation study design, with purposeful increases in intensity of implementation strategies. Each phase followed a standard operating procedure for all nursing units. The aims of our evaluation were to compare the following outcomes across phases:

- Implementation fidelity, adoption, cost, and nurse stakeholder experiences
- Impact of the implementation program on smoker engagement with the mHealth intervention

TIPS had three phases: two active implementation phases (15 weeks each) and one phase of sustainability monitoring (30 weeks). Clinical units received a different intensity of implementation strategies in the two active implementation phases (passive motivational posters only and posters + active nurse-led facilitation) to determine potential benefits of additional implementation strategies. We hypothesized that the active nurse-led inpatient strategy would result in higher postdischarge smoker engagement with the mHealth intervention compared with posters only. These two active 15-week phases were followed by a sustainability phase. To achieve our aims, we used mixed methods to determine the feasibility of the study design, implementation program, and patient engagement. These data were collected in preparation for a larger implementation trial. The study was approved by the University of Massachusetts Medical School institutional review board.

Implementation Framework

The implementation program design was guided by the Practical and Robust Implementation Science Model (PRISM) [14]. We selected PRISM out of many emerging implementation science theoretical frameworks because of its particular focus on challenges of evidence-based practice integration in the clinical setting and its integration of the diffusion of innovations theory to guide the uptake of technology-assisted interventions. PRISM guided our formative work and our effectiveness assessment using the reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) framework [15].

Setting and Sample

Four hospital units were recruited from two Northeast sites in the same network. We asked hospital leadership representatives to identify units containing high proportions of patients with chronic cardiac and pulmonary conditions. Since tobacco use heavily affects these prevalent conditions, the program could greatly benefit this population [16,17]. Nurse unit manager buy-in was obtained through meetings in which a nurse facilitator presented the program, assessed stakeholders' organizational perspectives as outlined by the PRISM model [14], and constructed a plan for communication between the facilitator and stakeholders during implementation.

Effectiveness Data-Supported Intervention

The TIPS implementation program supports the use of the effectiveness data-supported mHealth technology intervention: the texting system. Over the course of two weeks, the system delivered motivational messages written by smokers for smokers, encouraging participants to abstain from smoking [18]. These messages were created using current clinical guidelines [19] and social cognitive theory [20] and evaluated in the outpatient setting through a Web-assisted tobacco intervention [13,21-23]. This outpatient evaluation, a multisite randomized controlled trial with 900 smokers, found that motivational messages increased 6-month smoking cessation outcomes (odds ratio 1.7; 95% CI 1.0-2.8) compared with controls [13]. For our study, messages not appropriate for use in the inpatient setting (ie, "take a walk outside") were removed from the message protocol.

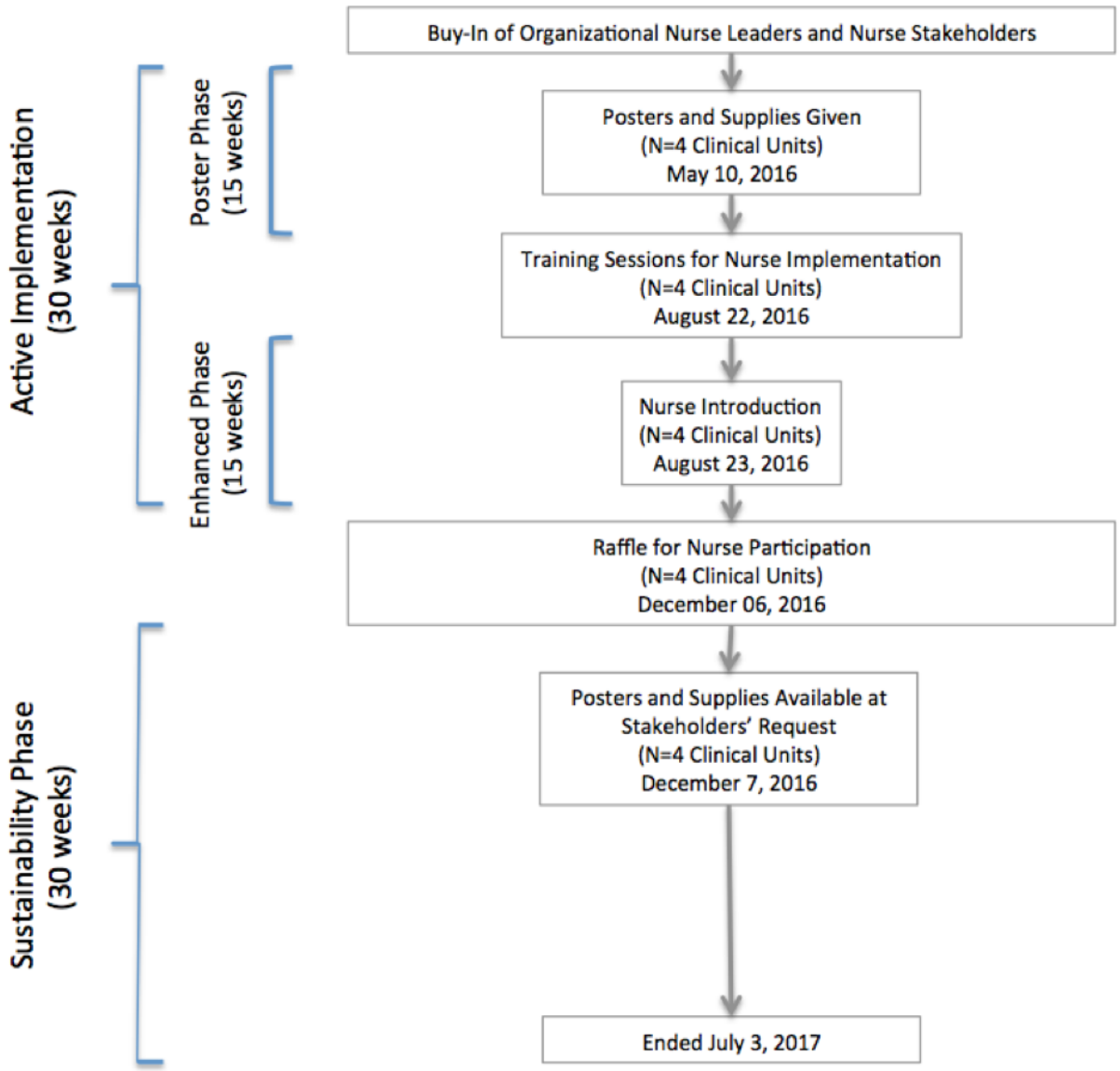
Implementation Program

TIPS included 30 weeks of active implementation and 30 weeks of sustainability monitoring for a total 60-week duration (Figure 1). Stakeholders including leaders, managers, and staff participated in a formative assessment leading to the final design of the TIPS program [24]. Weekly contact with nurse managers allowed feedback collection. Shared decision-making for adaptations to program strategies were elicited and typically implemented the following week.

Poster Phase

Active implementation included two phases of varying intensities. An initial 15-week phase, the poster phase, reflected the lowest program. All clinical units received promotional posters for use at the discretion of nurse unit managers. Poster content was created using health belief model concepts [24,25]. Posters were designed to motivate smokers to sign up for the texting system using their mobile phones. Communications and material safety hospital boards approved the posters. Managers received an informational letter containing guideline-based instructions for secure posting and optimal placement and additional materials to hang posters as requested throughout the study.

Figure 1. Study flow diagram.



At the end of the poster phase (15 weeks), two training sessions for nurses on each unit were performed by external facilitators (nurse scientists with clinical nurse specialist certification) to explain the TIPS program to staff nurses. The nurse training sessions were adapted with permission from a previously evaluated nurse education session for tobacco cessation [26]. Through nurse stakeholder feedback, education sessions were shortened to less than 10 minutes and presented during existing staff training times. Thereafter, nurse unit managers continued to train their staff.

Enhanced Phase

A second 15-week phase, the enhanced phase, used additional strategies for nurse-driven facilitation of the texting system including a protocol for introducing mobile messages and cue cards with the protocol strategically placed on nurse-specific computer carts. The protocol was shortened to three steps for ease of implementation: (1) ask on admission if the patient is a smoker, (2) point to the poster and give the patient an invitation to sign up, and (3) document the patient's decision in the electronic health record (EHR).

Sustainability Phase

In the 30-week sustainability phase, stakeholders were encouraged to contact the external facilitators to request supplies or communicate needs that arose. The promotional posters remained on patient room walls during the entire study.

Data Collection and Measures

Aim 1. Measures of Implementation Program Evaluation at the Unit and Nurse Level

Characteristics of units were collected from nurse unit managers; these included patient diagnoses and number of beds on the unit. We surveyed nurse unit managers' perceptions of readiness using the Organizational Readiness for Change Assessment (ORCA) survey [27], measuring three core elements: quality of evidence, environment or context for implementation, and facilitation of the implementation process.

Implementation fidelity, an agents' fidelity to the various elements of an intervention's protocol, includes the consistency of delivery of program components [15]. Process evaluation measures were identified from a taxonomy developed by Proctor and colleagues [28] and from a nurse-driven tobacco cessation intervention in the hospital setting by Duffy and colleagues [26]. Poster display fidelity was assessed weekly during active implementation and once during and after the sustainability phase. Poster fidelity was defined as the rate of missing posters by unit at each time point. Key dates when nurse managers facilitated the adopted implementation strategies were recorded for each unit; this reflected other milestone-driven implementation evaluation models [29]. Measuring implementation processes also involves identifying various changes made to strategies [30]. All adaptations to the program based on weekly stakeholder feedback were recorded. Fixed costs and variable costs were summed from administrative documents using a micro-costing system and reported by phase [31]. Interviews with stakeholders identified associated costs beyond these.

Barriers to patient adoption of the mHealth program were collected before and after active implementation by survey of the nurse staff. The question posed to nurses was, "Do you think any of the below reasons could be a barrier to introducing mobile messages to patients?" Responses were a checklist of potential barriers from a previously developed questionnaire [32] and those identified during the formative assessment [24].

Formative evaluation data on the strategies and intervention from purposive samples of nurses were collected by qualitative interview. Interview guides were developed using questions from previous evaluations [32]. We gathered acceptability data, asking the extent to which they agreed with the statement, "I would recommend other health care professionals to introduce these text messages as well."

Aim 2. Measures of Impact of the Implementation Program on Smoker Level Outcomes

Reach refers to "the absolute number, proportion, and representativeness of individuals who are willing to participate in a given initiative" [15]. This was reported as the absolute number of participants who signed up by week over the three phases as recorded in the mobile message system database. The number of smokers admitted to the units was measured during active implementation using EHR database reports designed for tobacco treatment specialist use; these listed current everyday and someday smokers by unit. To assist in examining reach by unit, the mobile message database captured date and content of received messages from smokers, such as self-reported location in the hospital.

Engagement has been used as a proximal measure for behavior change in previously tested digital behavior change interventions [33]. We measured daily engagement using the mobile message system database, with early disengagement of participants indicated by a response text message of "Stop." In addition to our quantitative outcome, we conducted follow-up interviews with smokers. In parallel to the nurse interviews, guides were developed from previous evaluations [32]. Acceptability was ascertained by asking smokers about their level of agreement with the statement, "I would recommend these text messages for others" [34].

Statistical Analyses

Implementation Program Evaluation at the Unit and Nurse Level

ORCA scores were calculated using a validated procedure, with mean scores reported for each survey category [27]. Poster fidelity was calculated as the rate of missing posters on patient walls by unit (the number of posters missing on units divided by the number of beds on the unit) for each time point. The average poster fidelity was calculated per unit for each phase during active implementation and at one time point during and one time point after the sustainability phase.

Nurse Experience With the Program

Perceived and actual barriers to patient participation in the intervention were reported as the percentage of nurses identifying each potential or observed barrier. Participants in TIPS were invited to give feedback on the program at the end

of the enhanced phase. Interviews were audio-recorded and transcribed. Interviews were first analyzed using a rapid identification of themes using audio recordings procedure to ascertain feedback needing immediate attention or action [35]. After this, transcripts were analyzed using open-ended coding: researchers reviewed coding together and agreed on the final data display in table format.

Impact of the Implementation Program on Smoker Level Outcomes

The percentage of smokers reached during active implementation was determined by the percentage of smokers who signed up for the intervention out of all smokers admitted to the units during the study period. Data from implementation process measures were juxtaposed with the absolute reach data to identify explanations for variation in reach over time. Using self-reported location data, the reach of participants was also compared between units by phase. The percentages of participants who sent stop messages, reported locations, and replied to a request for feedback were reported. Qualitative analytic procedures for smoker interviews were the same as the nurse interviews. Statistical tests were calculated using STATA 12.1 (StataCorp LLC) and qualitative analyses using NVivo 11.4 (QSR International).

Results

Aim 1. Implementation Program Evaluation at the Unit and Nurse Level

Characteristics of Units

Four hospital units were identified by organizational nurse leaders (Multimedia Appendix 1, Section A). Organizational readiness scores were similar, with each unit scoring high on the ORCA evidence, context, and facilitation scales.

Implementation Fidelity and Adaptability

The length of time for each unit to adopt the posters varied (Multimedia Appendix 1, Section B). In the poster phase, units had an average of 9.6% of posters missing from their walls, with a wide variation between units (range 1.7% to 16.2%). The overall rate of posters missing dropped in the enhanced phase to 1.9%, with a smaller range of missing posters (0.4% to 5.5%) per week. In the sustainability phase, the majority of units (3/4) had a lower rate of missing posters than during the initial poster phase.

Two units hung the posters immediately in week 1, while the other units waited until week 3 and week 5. Stakeholders noted that competing unit priorities and lack of available personnel created barriers to adoption. Stakeholders decided education sessions were to be held during preexisting nurse staff huddles in a common work area on the units. Nurse managers independently facilitated education sessions for their staff nurses: units in one hospital site (units 2 and 4) independently facilitated 7 and 8 additional sessions each, and units in the other site (units 1 and 3) independently facilitated two each.

Unexpected barriers to poster implementation arose, including rooms with contact precautions and wall painting. To overcome these barriers, nurse managers requested business cards containing the poster graphic; these cards were then inserted into standard informational packets given to patients upon admission. In the enhanced phase, manager feedback was integral in creating a visual cue, or cue card, reminding nurses to introduce the intervention and follow the 3-step protocol. Stakeholders identified a need for more nurse engagement beyond training (Multimedia Appendix 2). Tent cards to reiterate training (week 20), paper surveys to counteract low institutional email use (week 25), and feedback boards displaying the unit's success in introducing the intervention to patients (week 28) were all placed in nurse break rooms. Nurse unit managers did not suggest further adaptations during the sustainability phase.

Totaling just US $2925 overall, this implementation program was very low cost (Multimedia Appendix 3). Even with potential time cost, stakeholders did not identify any additional costs related to the program or implementation beyond materials supplied to the unit. Initial fixed costs of laminated posters and affixation materials were more than three-quarters (US $2240, 76.6%) of the total cost; maintenance costs were negligible. After initial fixed costs were spent, the 60-week program cost US $6.63 (US $683.28/103) per smoker reached and US $8.33 per smoker engaged (US $683.28/82).

Nurse Experience With the Program

Nurse survey data, collected at baseline (just after training) and follow-up (at the end of active implementation), showed that nurses overestimated patient and technology barriers to smokers signing up for the intervention compared with nurse report at the end of the enhanced phase (Table 1). One barrier, patient motivation, was severely underestimated (anticipated: n=8, 26%; actual: n=17, 53%).

Program feedback is reported (Multimedia Appendix 4). The number of patients that nurses introduced the program to varied widely, from 1 to 20 people (median 3). Acceptability of the program was indicated, as nurses either strongly agreed or agreed that they would recommend the program to other health care professionals. A perception emerged that the program was a success.

> *I think that it went very well, that the staff were much more engaged than I thought they were gonna be and that it went really well.* [Nurse manager]

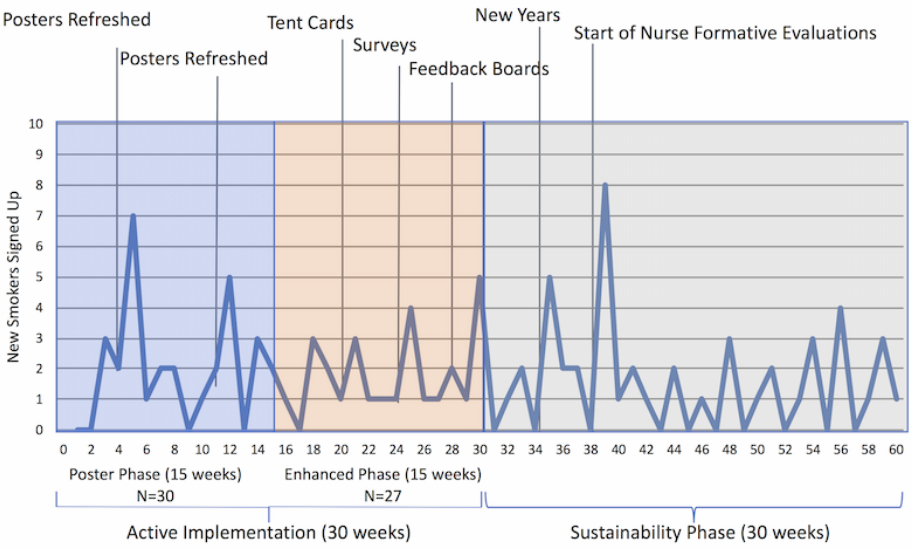
Table 1. Perceived and actual barriers to patient reach by nurses in the enhanced phase.

Barriers to patient reach[a]	Anticipated barriers at baseline (n=31), n (%)	Reported barriers at follow-up (n=32), n (%)	Percentage difference[b]
Patient characteristics			
Language barrier	15 (48)	5 (16)	–32.8
Patient cognition barrier	13 (42)	3 (9)	–32.5
Older age of patient	6 (19)	3 (9)	–9.9
Patient on other substances and a barrier to interest in quitting tobacco	8 (26)	8 (25)	–0.8
Technology issues			
Patient does not have phone	10 (32)	5 (16)	–16.7
Patient concerned about charges for text messages	8 (26)	3 (9)	–16.4
Patient does not text	6 (19)	4 (13)	–6.8
Patient left phone at home	5 (16)	3 (9)	–6.7
Motivation			
Patient not motivated to sign up	8 (26)	17 (53)	27.3

[a]Barriers are organized by largest differences in percentages by category.

[b]Percentage of barriers anticipated at baseline subtracted by the percentage of barriers reported at follow-up.

Figure 2. Reach of the Technology Inpatient Program for Smokers intervention by week and phase.



Aim 2. Impact of the Implementation Program on Smoker Level Outcomes

A total of 103 smokers signed up for the texting intervention over a 60-week period, with varying implementation strategy intensity in all 3 phases (Figure 2). During active implementation, 57 smokers signed up out of a potential 783 smokers admitted to the units (7.3% reach). In the initial poster phase, lasting 15 weeks, 30 smokers signed up. There were large fluctuations by week, and some of the fluctuations mirrored the times when nurse managers took action to facilitate the intervention. For example, during week 4, a unit refreshed their posters, and in week 5, another unit hung their posters for the first time, with a subsequent spike in reach of smokers following these events. In all, 7.1% (30/421) of patients identified as smokers in the EHR signed up for the program (30/328, 9.1% of everyday smokers).

During the nurse staff–facilitated 15-week enhanced phase, 27 smokers signed up. Several peaks in reach during weeks 21, 25, and 30 followed new implementation strategies. Consistent sign-up, with at least one smoker in 14 of the 15 weeks, improved upon the poster phase (11/15 weeks). A total of 7.5% (27/362) of patients identified as smokers in the EHR signed up for the program (27/285, 9.5% of everyday smokers). Throughout the 30-week sustainability phase, 46 smokers signed up, with sign-up steadily decreasing over time (weeks 31-45: mean 1.7, weeks 46-60: mean 1.3). Reach increased in weeks following the start of the New Year (week 34) and following staff interviews (week 38).

Reach of the technology by site was also examined. Analysis was performed with smokers who reported their location (46/103, 44.7%) and excluded several participants not on a hospital unit (3/103) or who were unsure what unit they were on (2/103), for a total of 39.8% (41/103) responses used (Figure 3). In the enhanced phase, one site (unit 1) accounted for over half (8/14, 57.1%) of smokers enrolled, whereas an alternate unit (unit 4) enrolled over half of smokers (6/11, 54.5%) in the first half of the sustainability phase as well (weeks 31-45).

Overall, 79.6% (82/103) of participants completed the full 2-week duration of the intervention (Table 2). Engagement increased from the poster phase (22/30, 73.3%) to the enhanced phase (23/27, 85.2%) when nurse facilitation was initiated. A similar pattern occurred for smoker response to two-way text questions. Responsiveness to patient location more than doubled from poster phase to enhanced phase (2.2, 59.3/26.7) and patient responsiveness to an inquiry for feedback almost quadrupled (3.9, 65.2/16.7). For all measures, engagement declined steadily in the sustainability phase but did not drop to initial poster phase levels.

In follow-up interviews with smokers, text messages were described as easy to read and understandable to all smokers (Multimedia Appendix 5). The majority of interviewed smokers reported that the program made them more serious about quitting. Two-thirds of smokers interviewed reduced their cigarette consumption considerably while using the mHealth intervention, with 2 smokers quitting entirely during and following intervention use. Two smokers were signed up by a family member, a patient was signed up by her sister, and a visitor was signed up by her significant other. The latter two both reported that their family member took the smoker's phone and texted the number on the poster. Acceptability of the text messages was indicated as smokers strongly agreed or agreed that they would recommend the text messages to others.

Figure 3. Reach of the Technology Inpatient Program for Smokers intervention by unit and phase. Analysis included only participants who reported location (N=41). Missing responses due to no response (N=57), smokers not on a hospital unit (N=3), and smokers who were unsure what unit they were on (N=2) are assumed to be at random. Mean new smokers signed up for all units over all phases is 2.6 smokers, represented by a dashed line.

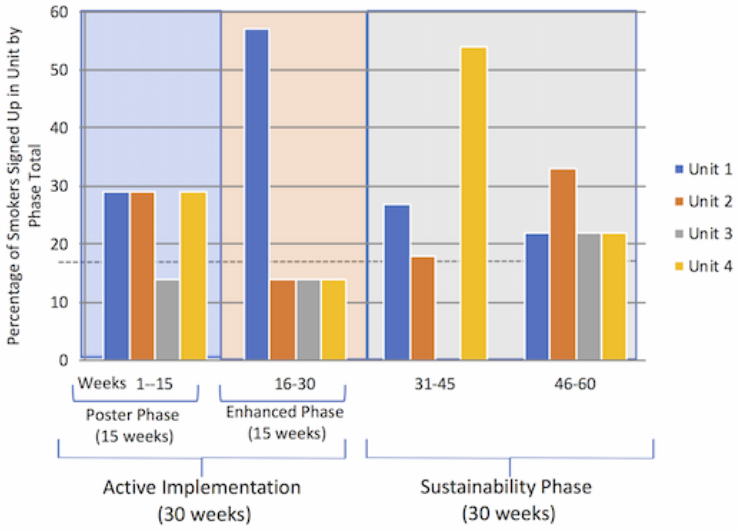


Table 2. Technology Inpatient Program for Smokers mHealth intervention engagement (main outcome) by smokers.

Phase and total reach	Weeks	Smoker response to two-way text questions during 2-week mHealth intervention		Main outcome (smoker engagement in the mHealth intervention)	
		Responded with location, n (%)	Responded to feedback inquiry, n (%)	Completed full 2-week mHealth intervention, n (%)	Disengaged (did not complete full 2-week mHealth intervention), n (%)
Poster (n=30)	1-15	8 (27)	4 (17)	22 (73)	8 (27)
Enhanced (n=27)	16-30	16 (59)	15 (65)	23 (85)	4 (15)
Sustainability (n=26)	31-45	13 (50)	14 (58)	22 (85)	5 (15)
Sustainability (n=20)	46-60	9 (45)	6 (35)	15 (75)	5 (25)
Total (N=103)	N/A[a]	46 (45)	39 (44)	82 (80)	21 (20)

[a]Not applicable.

Discussion

Principal Findings

We successfully completed a two-site implementation of the TIPS program. This program was readily sustained over 60 weeks. To our knowledge, our mHealth intervention is the first Quit Smoking texting system to be offered to patients in hospital units for continued use after discharge. Our findings support the feasibility of engaging smokers to adopt a texting system during the forced abstinence of their inpatient stay. As a light-touch, low-maintenance implementation program, TIPS resulted in an average of 1.7 smokers engaged in the texting system every week. Unexpectedly, nurse-facilitated delivery of the intervention during the enhanced phase did not lead to an increase in the number of smokers adopting the text message intervention. However, as hypothesized, of the smokers who did adopt during this phase, the majority sustained 2-week engagement and at a higher rate than those who adopted during the poster phase.

Adoption of TIPS was sustained over time. While implementation fatigue is often a barrier to consistent evidence-based practice implementation in hospital settings [36], TIPS was well sustained. Organizational theory supports middle managers as key to staff engagement [37] and quality improvement integration within the hospital setting [38]. Managers were able to make changes on their floors through simple educational and maintenance processes. This approach was evidently well received by managers, who participated enthusiastically in new strategy creation, took ownership of hanging posters, and undertook externally implemented staff education. While managers reported the program engaged nurses more than other quality improvement initiatives, excitement about the program faded when new adaptations to strategies were not being implemented.

> *I think they were pretty good about buying in, but it was kind of there was excitement, and then they lost it. There'd be an excitement again, and then they'd lose it, which if we had that magic wand to keep the excitement going, it'd be great.* [Nurse manager]

These findings suggest that continual reevaluation of implementation efforts and tailoring those efforts to the unique cultures and practices of different units is a key to sustainability.

Comparison With Prior Results

Over 60 weeks, a total of 103 smokers adopted the mHealth intervention. Implementation fidelity of posters in the patient rooms was a predictor of reach over time. While posters combined with health care staff training have been used in the implementation of evidence-based practices in hospitals in previous studies [39-43], the relationship between fidelity and reach has been largely unexamined. Our findings were consistent with an mHealth tobacco cessation intervention (iQuit) implemented outside of the hospital setting using an intensive recruitment strategy [44]. Still, a minority of smokers admitted to these units during active implementation, identified using the EHR, adopted the text messaging program. Accuracy of the records, exposure of patients who smoked to posters, and patients' physical or cognitive ability to sign up while hospitalized are unknown using this low-intensity method. Nevertheless, a minority of the smokers who likely could have adopted did. Nurses reported twice the rate of nonmotivated smokers than anticipated.

> *I think the challenge is getting people to actually want to sign up. But the actual sign-up process is pretty simple.* [Nurse]

While the hospital is a teachable moment for some smokers [45], others may need additional help getting motivated. Additional content for these motivational phase smokers should be created and tested.

Engagement with the behavioral intervention is a critical component in the efficacy of an intervention, ensuring smokers receive the full benefit of the 2-week intervention [46]. Overall engagement with text messages in TIPS was similar to prior studies, such as iQuit, where 81% of smokers completed an mHealth intervention [44]. TIPS improved upon longer duration studies in which 31.8% [47] and 45% [48] stopped messages early. Smokers may not have had a fully framed understanding of what the messaging entailed with poster-only facilitation, while nurse delivery allowed for clarification and use of behavior change techniques like persuasive argument, which may have led to stronger engagement with the messages [49]. In her behavior change technique taxonomy, Susan Mitchie and colleagues [49] identified core techniques across theoretical frameworks for behavior change, which include persuasive argument, health consequences, and action planning. Physicians and nurses have used these techniques in tobacco cessation interventions in the past, showing an increased likelihood for cessation [45,50]. We highlighted these techniques during short unit-level nurse training sessions for staff.

Prior work has shown that hospitalization is a teachable moment for families as well as patients, providing enhanced motivation to quit or to stimulate quitting attempts [51]. We similarly found the hospital setting to be a teachable moment for visitors of patients. A surprising finding during formative evaluation of smoker experiences was talking to family members of patients who assisted them in signing up or signed up themselves. Thus far, tobacco-using parents of newborns and hospitalized children are the only populations of family members who have been reached with tobacco cessation interventions in the hospital setting [52-54]. Engaging family members of hospitalized adults in tobacco cessation and considering the role of family or visitors as an avenue of reaching hospitalized smokers are gaps in current tobacco cessation interventions which our implementation strategy might be poised to surmount.

Limitations

There are limitations to our study. A single health system is not fully generalizable to other inpatient settings, although we did see success in units of diverse specialties, structures, and characteristics across two hospital sites. Six months of active implementation is short for staff practice change, yet nurse managers saw exceptionally high staff engagement during that time period. We collected limited smoker data in an effort to avoid burdening smokers and impeding reach. Anonymity may

have been a perceived benefit of mobile messages that helped drive participation. For smokers who disengaged early from the intervention, indicating they would like to stop receiving messages, we did not reach out to determine why. While numerous nurses did introduce the program to smokers, we did not ascertain how many smokers were successfully reached and engaged through posters alone versus posters plus nurse facilitation in the enhanced phase [55]. Process measures for nurse implementation, beyond a final count of smokers reached, may need to be developed further determine the pathway between strategy and smoker engagement.

Conclusions

TIPS is a low-intensity, sustained program engaging inpatient smokers. While our intervention reached a minority of admitted smokers, the results were comparable to intensive and costly intervention strategies in the outpatient setting [44]. As overall rates of smoking decline, smokers become increasingly challenging to reach and engage. Nurses reported that half of the smokers they approached were not motivated to quit, highlighting the necessity to infuse materials and nurse training with motivational phase-specific strategies. Interestingly, patient family members helped smokers adopt the intervention and even signed up themselves, suggesting an intriguing new public health strategy for using the hospital setting to teach visitors as well as patients. Smokers cluster in social networks, so interventions that can reach and engage both the patient and family may be an exciting innovation. Continued testing of strategies to sustain nurse engagement in facilitation of evidence-based interventions is needed. TIPS represents an innovative, low-cost, easily disseminated strategy for engaging nurses and reaching patients with behavior change interventions.

Acknowledgments

We would like to thank the nurses who gave of their time and expertise to inform and evaluate the program. The work you do every day is inspiring. Funding for the study was received from the National Cancer Institute (R01 CA190866-01A1) and the University of Massachusetts Medical School's Center for Clinical and Translational Science (1UL1RR031982-01 U54). ACB is supported by the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Health Services Research, the VA Center for Healthcare Organizational and Implementation Research, the VA Center for Clinical Management Research, and the University of Massachusetts Medical School's Prevention and Control of Cancer: Post-doctoral Training in Implementation Science Program (1R25CA172009-02). RSS is funded by a National Cancer Institute Career Development Award (K07 CA172677).

Authors' Contributions

ACB developed the implementation program; carried out implementation efforts; collected, analyzed and interpreted the data; and drafted the manuscript. ND made substantial contributions to design of materials and acquisition of data. AP made substantial contributions to analysis and interpretation of data. TKH developed and tested the behavioral texting intervention in previous work. TPH, RSS, and TKH made substantial contributions to the conception, interpretation of analysis and findings, and critical revision for intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Technology Inpatient Program for Smokers adoption and implementation by unit and phase.
[PDF File (Adobe PDF File)62 KB - mhealth_v7i10e14331_app1.pdf]

Multimedia Appendix 2
Materials to engage nurses in program facilitation.
[PNG File 278 KB - mhealth_v7i10e14331_app2.png]

Multimedia Appendix 3
Program cost evaluation.
[PDF File (Adobe PDF File)47 KB - mhealth_v7i10e14331_app3.pdf]

Multimedia Appendix 4
Themes from qualitative analysis of program feedback interviews with nurses.
[PDF File (Adobe PDF File)56 KB - mhealth_v7i10e14331_app4.pdf]

Multimedia Appendix 5
Themes from qualitative analysis of program feedback interviews with smokers.

[PDF File (Adobe PDF File)52 KB - mhealth_v7i10e14331_app5.pdf]

References

1. National Center for Chronic Disease Prevention and Health Promotion. The health consequences of smoking—50 years of progress. A report of the Surgeon General. Atlanta: Centers for Disease Control and Prevention; 2014. URL: https://www.ncbi.nlm.nih.gov/books/n/surgsmoke50/pdf/ [accessed 2019-09-24]
2. Xu X, Bishop EE, Kennedy SM, Simpson SA, Pechacek TF. Annual healthcare spending attributable to cigarette smoking: an update. Am J Prev Med 2015 Mar;48(3):326-333 [FREE Full text] [doi: 10.1016/j.amepre.2014.10.012] [Medline: 25498551]
3. Hasdai D, Garratt KN, Grill DE, Lerman A, Holmes DR. Effect of smoking status on the long-term outcome after successful percutaneous coronary revascularization. N Engl J Med 1997 Mar 13;336(11):755-761. [doi: 10.1056/NEJM199703133361103] [Medline: 9052653]
4. Borrelli B, Bartlett YK, Tooley E, Armitage CJ, Wearden A. Prevalence and frequency of mhealth and ehealth use among US and UK smokers and differences by motivation to quit. J Med Internet Res 2015 Jul 04;17(7):e164 [FREE Full text] [doi: 10.2196/jmir.4420] [Medline: 26149323]
5. Bergvik S, Wynn R. The use of short message service (SMS) among hospitalized coronary patients. Gen Hosp Psychiatry 2012;34(4):390-397. [doi: 10.1016/j.genhosppsych.2012.03.006] [Medline: 22516220]
6. US Department of Health and Human Services. 2014. Tobacco use and secondhand smoke exposure: mobile phone-based cessation interventions URL: https://www.thecommunityguide.org/sites/default/files/assets/Tobacco-Mobile-Phone-based-Cessation.pdf [accessed 2019-09-12]
7. Rigotti NA, Munafo MR, Stead LF. Interventions for smoking cessation in hospitalised patients. Cochrane Database Syst Rev 2007 Jul 18(3):CD001837. [doi: 10.1002/14651858.CD001837.pub2] [Medline: 17636688]
8. Campbell S, Pieters K, Mullen K, Reece R, Reid RD. Examining sustainability in a hospital setting: case of smoking cessation. Implement Sci 2011;6:108 [FREE Full text] [doi: 10.1186/1748-5908-6-108] [Medline: 21917156]
9. Duffy SA, Ronis DL, Karvonen-Gutierrez CA, Ewing LA, Dalack GW, Smith PM, et al. Effectiveness of the tobacco tactics program in the Department of Veterans Affairs. Ann Behav Med 2014 Oct;48(2):265-274. [doi: 10.1007/s12160-014-9605-z] [Medline: 24823842]
10. Freund M, Campbell E, Paul C, Sakrouge R, Wiggers J. Smoking care provision in smoke-free hospitals in Australia. Prev Med 2005 Jul;41(1):151-158. [doi: 10.1016/j.ypmed.2004.09.044] [Medline: 15917006]
11. Li I, Lee SD, Chen C, Jeng Y, Chen Y. Facilitators and barriers to effective smoking cessation: counselling services for inpatients from nurse-counsellors' perspectives—a qualitative study. Int J Environ Res Public Health 2014 May 06;11(5):4782-4798 [FREE Full text] [doi: 10.3390/ijerph110504782] [Medline: 24806190]
12. Mazza R, Lina M, Invernizzi G, Pierotti M, De Marco C, Borreani C, et al. The gap between tobacco treatment guidelines, health service organization, and clinical practice in comprehensive cancer centres. J Oncol 2011;2011:145617 [FREE Full text] [doi: 10.1155/2011/145617] [Medline: 21776269]
13. Houston TK, Sadasivam RS, Allison JJ, Ash AS, Ray MN, English TM, et al. Evaluating the QUIT-PRIMO clinical practice ePortal to increase smoker engagement with online cessation interventions: a national hybrid type 2 implementation study. Implement Sci 2015;10:154 [FREE Full text] [doi: 10.1186/s13012-015-0336-8] [Medline: 26525410]
14. Feldstein AC, Glasgow RE. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. Jt Comm J Qual Patient Saf 2008 Apr;34(4):228-243. [Medline: 18468362]
15. Reach, effectiveness, adoption, implementation, maintenance. URL: http://re-aim.org [accessed 2019-09-12]
16. Pfuntner A, Wier L, Stocks C. Most frequent conditions in US hospitals, 2010. HCUP Statistical Brief #148 2013 Jan. [doi: 10.1111/j.1746-9066.2013.00026.x] [Medline: 23534077]
17. United States Public Health Service. How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: a report of the Surgeon General. Atlanta: US Department of Health and Human Services; 2010. URL: https://www.ncbi.nlm.nih.gov/books/n/rptsmokedis/pdf/ [accessed 2019-09-24]
18. Sadasivam RS, Volz EM, Kinney RL, Rao SR, Houston TK. Share2Quit: web-based peer-driven referrals for smoking cessation. JMIR Res Protoc 2013;2(2):e37 [FREE Full text] [doi: 10.2196/resprot.2786] [Medline: 24067329]
19. Fiore C, Jaen CR, Baker TB. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A US Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
20. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs: Prentice-Hall, Inc; 1986.
21. Houston TK, Sadasivam RS, Ford DE, Richman J, Ray MN, Allison JJ. The QUIT-PRIMO provider-patient Internet-delivered smoking cessation referral intervention: a cluster-randomized comparative effectiveness trial: study protocol. Implement Sci 2010 Nov 17;5:87 [FREE Full text] [doi: 10.1186/1748-5908-5-87] [Medline: 21080972]
22. Houston TK, Ford DE. A tailored Internet-delivered intervention for smoking cessation designed to encourage social support and treatment seeking: usability testing and user tracing. Inform Health Soc Care 2008 Mar;33(1):5-19. [doi: 10.1080/14639230701842240] [Medline: 18604759]

23. Sadasivam RS, Delaughter K, Crenshaw K, Sobko HJ, Williams JH, Coley HL, et al. Development of an interactive, Web-delivered system to increase provider-patient engagement in smoking cessation. J Med Internet Res 2011 Oct 18;13(4):e87 [FREE Full text] [doi: 10.2196/jmir.1721] [Medline: 22011394]
24. Blok A, Sadasivam R, Houston T. Formative evaluation to determine facilitators and barriers to nurse-driven implementation: designing an inpatient mhealth intervention to support smoking cessation. 2016 Presented at: NAME; 2016; LOCATION. [doi: 10.24251/hicss.2017.401]
25. Rimer B, Glanz K. Theory at a Glance: A Guide for Health Promotion Practice. Bethesda: National Institutes of Health; 2005.
26. Duffy SA, Ronis DL, Titler MG, Blow FC, Jordan N, Thomas PL, et al. Dissemination of the nurse-administered Tobacco Tactics intervention versus usual care in six Trinity community hospitals: study protocol for a comparative effectiveness trial. Trials 2012 Aug 01;13:125 [FREE Full text] [doi: 10.1186/1745-6215-13-125] [Medline: 22852834]
27. Helfrich CD, Li Y, Sharp ND, Sales AE. Organizational readiness to change assessment (ORCA): development of an instrument based on the Promoting Action on Research in Health Services (PARIHS) framework. Implement Sci 2009 Jul 14;4:38 [FREE Full text] [doi: 10.1186/1748-5908-4-38] [Medline: 19594942]
28. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
29. Chamberlain P, Brown CH, Saldana L. Observational measure of implementation progress in community based settings: the Stages of Implementation Completion (SIC). Implement Sci 2011 Oct 06;6:116 [FREE Full text] [doi: 10.1186/1748-5908-6-116] [Medline: 21974914]
30. Brownson RC, Colditz GA, Proctor EK, editors. Dissemination and Implementation Research in Health: Translating Science to Practice. Oxford: Oxford University Press; 2018.
31. Frick KD. Microcosting quantity data collection methods. Med Care 2009 Jul;47(7 Suppl 1):S76-S81 [FREE Full text] [doi: 10.1097/MLR.0b013e31819bc064] [Medline: 19536026]
32. Duffy SA, Ronis DL, Ewing LA, Waltje AH, Hall SV, Thomas PL, et al. Implementation of the Tobacco Tactics intervention versus usual care in Trinity Health community hospitals. Implement Sci 2016 Nov 04;11(1):147 [FREE Full text] [doi: 10.1186/s13012-016-0511-6] [Medline: 27814722]
33. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
34. Blok AC, Blonquist TM, Nayak MM, Somayaji D, Crouter SE, Hayman LL, et al. Feasibility and acceptability of "healthy directions," a lifestyle intervention for adults with lung cancer. Psychooncology 2018 Jan;27(1):250-257 [FREE Full text] [doi: 10.1002/pon.4443] [Medline: 28426922]
35. Neal JW, Neal ZP, VanDyke E, Kornbluh M. Expediting the analysis of qualitative data in evaluation. Am J Eval 2014 May 27;36(1):118-132. [doi: 10.1177/1098214014536601]
36. Huynh C, Bowles D, Yen M, Phillips A, Waller R, Hall L, et al. Change implementation: the association of adaptive reserve and burnout among inpatient medicine physicians and nurses. J Interprofessional Care 2018 Mar 20;32(5):549-555. [doi: 10.1080/13561820.2018.1451307]
37. Engle RL, Lopez ER, Gormley KE, Chan JA, Charns MP, Lukas CV. What roles do middle managers play in implementation of innovative practices? Health Care Manage Rev 2017;42(1):14-27 [FREE Full text] [doi: 10.1097/HMR.0000000000000090] [Medline: 26488239]
38. White M, Butterworth T, Wells JS. Healthcare quality improvement and 'work engagement'; concluding results from a national, longitudinal, cross-sectional study of the 'Productive Ward-Releasing Time to Care' Programme. BMC Health Serv Res 2017 Aug 01;17(1):510 [FREE Full text] [doi: 10.1186/s12913-017-2446-2] [Medline: 28764696]
39. Gupta AK, Gupta AC, Gupta A, Ranga SS, Rewari BB, Bansal AP. Implementation and impact of a post-exposure prophylaxis helpline. Occup Med (Lond) 2015 Jul;65(5):398-401. [doi: 10.1093/occmed/kqv053] [Medline: 25941211]
40. Kampf G, Reise G, James C, Gittelbauer K, Gosch J, Alpers B. Improving patient safety during insertion of peripheral venous catheters: an observational intervention study. GMS Hyg Infect Control 2013;8(2):Doc18 [FREE Full text] [doi: 10.3205/dgkh000218] [Medline: 24327944]
41. Sedlmayr B, Patapovas A, Kirchner M, Sonst A, Müller F, Pfistermeister B, et al. Comparative evaluation of different medication safety measures for the emergency department: physicians' usage and acceptance of training, poster, checklist and computerized decision support. BMC Med Inform Decis Mak 2013 Jul 29;13:79 [FREE Full text] [doi: 10.1186/1472-6947-13-79] [Medline: 23890121]
42. Lee YL, Yip WK, Goh BW, Chiam EP, Ng HP. Fall prevention among children in the presence of caregivers in a paediatric ward: a best practice implementation. Int J Evid Based Healthc 2013 Mar;11(1):33-38. [doi: 10.1111/1744-1609.12003] [Medline: 23448328]
43. Dale C, Fowler RA, Adhikari NKJ, Pinto R, Rose L. Implementation of a research awareness program in the critical care unit: effects on families and clinicians. Intensive Crit Care Nurs 2010 Apr;26(2):69-74. [doi: 10.1016/j.iccn.2009.09.003] [Medline: 19864137]

44. Naughton F, Jamison J, Boase S, Sloan M, Gilbert H, Prevost AT, et al. Randomized controlled trial to assess the short-term effectiveness of tailored web- and text-based facilitation of smoking cessation in primary care (iQuit in practice). Addiction 2014 Jul;109(7):1184-1193 [FREE Full text] [doi: 10.1111/add.12556] [Medline: 24661312]
45. Rigotti NA, Clair C, Munafò MR, Stead LF. Interventions for smoking cessation in hospitalised patients. Cochrane Database Syst Rev 2012;5:CD001837 [FREE Full text] [doi: 10.1002/14651858.CD001837.pub3] [Medline: 22592676]
46. Christofferson DE, Hertzberg JS, Beckham JC, Dennis PA, Hamlett-Berry K. Engagement and abstinence among users of a smoking cessation text message program for veterans. Addict Behav 2016 Nov;62:47-53. [doi: 10.1016/j.addbeh.2016.06.016] [Medline: 27318948]
47. Balmford J, Leifert JA, Schulz C, Elze M, Jaehne A. Implementation and effectiveness of a hospital smoking cessation service in Germany. Patient Educ Couns 2014 Jan;94(1):103-109. [doi: 10.1016/j.pec.2013.09.024] [Medline: 24176610]
48. Naughton F, Cooper S, Bowker K, Campbell K, Sutton S, Leonardi-Bee J, et al. Adaptation and uptake evaluation of an SMS text message smoking cessation programme (MiQuit) for use in antenatal care. BMJ Open 2015 Oct 22;5(10):e008871 [FREE Full text] [doi: 10.1136/bmjopen-2015-008871] [Medline: 26493459]
49. 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]
50. Rice VH, Heath L, Livingstone-Banks J, Hartmann-Boyce J. Nursing interventions for smoking cessation. Cochrane Database Syst Rev 2017 Dec 15;12:CD001188. [doi: 10.1002/14651858.CD001188.pub5] [Medline: 29243221]
51. Dohnke B, Ziemann C, Will KE, Weiss-Gerlach E, Spies CD. Do hospital treatments represent a 'teachable moment' for quitting smoking? A study from a stage-theoretical perspective. Psychol Health 2012;27(11):1291-1307. [doi: 10.1080/08870446.2012.672649] [Medline: 22472171]
52. Bock BC, Becker BM, Borrelli B. Smoking behavior and risk perception among the parents of infants in the neonatal intensive care unit. Nicotine Tob Res 2008 Jan;10(1):47-54. [doi: 10.1080/14622200701767795] [Medline: 18188744]
53. Phillips RM, Merritt TA, Goldstein MR, Deming DD, Slater LE, Angeles DM. Prevention of postpartum smoking relapse in mothers of infants in the neonatal intensive care unit. J Perinatol 2012 May;32(5):374-380 [FREE Full text] [doi: 10.1038/jp.2011.106] [Medline: 21836549]
54. Winickoff JP, Healey EA, Regan S, Park ER, Cole C, Friebely J, et al. Using the postpartum hospital stay to address mothers' and fathers' smoking: the NEWS study. Pediatrics 2010 Mar;125(3):518-525. [doi: 10.1542/peds.2009-0356] [Medline: 20123776]
55. Duffy SA, Ewing LA, Louzon SA, Ronis DL, Jordan N, Harrod M. Evaluation and costs of volunteer telephone cessation follow-up counseling for veteran smokers discharged from inpatient units: a quasi-experimental, mixed methods study. Tob Induc Dis 2015;13(1):4 [FREE Full text] [doi: 10.1186/s12971-015-0028-9] [Medline: 25674045]

Abbreviations

EHR: electronic health record
mHealth: mobile health
ORCA: Organizational Readiness for Change Assessment
PRISM: Practical and Robust Implementation Science Model
RE-AIM: reach, effectiveness, adoption, implementation, maintenance
TIPS: Technology Inpatient Program for Smokers

Edited by G Eysenbach; submitted 10.04.19; peer-reviewed by A Khoja, T Langley; comments to author 25.05.19; revised version received 19.07.19; accepted 21.07.19; published 04.10.19.

Please cite as:
Blok AC, Sadasivam RS, Hogan TP, Patterson A, Day N, Houston TK
Nurse-Driven mHealth Implementation Using the Technology Inpatient Program for Smokers (TIPS): Mixed Methods Study
JMIR Mhealth Uhealth 2019;7(10):e14331
URL: https://mhealth.jmir.org/2019/10/e14331
doi:10.2196/14331
PMID:31588908

Original Paper

Content and Feature Preferences for a Physical Activity App for Adults With Physical Disabilities: Focus Group Study

Sara H Olsen[1], MS, MPH; Sandra L Saperstein[1], PhD; Robert S Gold[1], PhD, DrPH

Department of Behavioral and Community Health, School of Public Health, University of Maryland, College Park, MD, United States

Corresponding Author:
Sara H Olsen, MS, MPH
Department of Behavioral and Community Health
School of Public Health
University of Maryland
4200 Valley Dr #1234
College Park, MD, 20742
United States
Phone: 1 301 405 2463
Email: sara.olsen@gmail.com

Abstract

Background: Hundreds of thousands of mobile phone apps intended to improve health and fitness are available for download across platforms and operating systems; however, few have been designed with people with physical disabilities in mind, ignoring a large population that may benefit from an effective tool to increase physical activity.

Objective: This study represents the first phase in the development process of a fitness tracking app for people with physical disabilities interested in nontraditional sport. The aim of this research was to explore user preferences for content, appearance, and operational features of a proposed physical activity app for people with physical disabilities to inform the design of a mobile phone app for increasing physical activity.

Methods: Four focus groups were conducted with 15 adults with physical disabilities who currently participate in nontraditional, non-Paralympic sport. Data collected from the focus group sessions centered on content, functionality, and appearance of apps currently used by participants as well as preferences for a future app.

Results: Participants (mean age 35.7, SD 9.2 years) were mostly white (13/15, 87%), and all were currently participating in CrossFit and at least one other sport. Five main themes were identified. Themes included preferences for (1) workout-specific features that were tailored or searchable by disability, (2) user experience that was intuitive and accessible, (3) profile personalization options, (4) gamification features that allowed for competition with self and other users, and (5) social features that allowed increased interaction among users. Participants expressed a primary interest in having a fitness app that was designed for people with physical disabilities such that the features present in other fitness tracking apps were relevant to them and their community of adaptive athletes.

Conclusions: The results showed that features related to user experience, social engagement, and gamification are considered important to people with physical disabilities. Features highlighted by participants as most desired, from a consumer perspective, were in line with research identifying attributes of quality apps that use behavior change techniques to influence positive physical activity behavior change. Such insights should inform the development of any fitness app designed to integrate users with disabilities as a primary user base.

(JMIR Mhealth Uhealth 2019;7(10):e15019) doi:10.2196/15019

KEYWORDS

adapted sport; disability; physical activity; focus group; apps; fitness; user-centered design; mHealth

Introduction

Background

The physical, psychological, emotional, and economic benefits of participating in physical activity are well established. Regular, moderate-intensity physical activity has been shown to improve muscular and cardiorespiratory fitness as well as bone health and reduce the risk of chronic diseases such as hypertension, coronary heart disease, diabetes, various cancers, and depression. Physical activity plays a critical role in weight control, which has also been linked to the same chronic diseases [1-4]. Conversely, physical inactivity has been attributed to 6% of coronary heart disease, 7% of type 2 diabetes, 10% of breast cancer, and 10% of colon cancer incidence globally [5]. Research also suggests that even small increases in physical activity in the least active populations have a much larger impact on overall health than any increase in activity level in populations closer to achieving physical activity guidelines [5]. These findings provide a powerful rationale for interventions to increase physical activity.

The Centers for Disease Control and Prevention broadly defines disability in adults as a limitation to routine activities or the need for assistive devices. More than half of the adult US population with any disability is inactive—47% are completely sedentary and an additional 22% do not move enough to meet the physical activity guidelines [6], which is higher than the national sedentary estimate of 23.7% in the general population [7]. This study focuses on people with physical disabilities, that is, those individuals with permanent physical impairments that affect participation in physical activity and fitness. People with physical disabilities, on the whole, are more likely to report poor overall health, lower access to health care, and decreased physical activity [8] than those without disabilities. This population is at significantly greater risk of comorbidities associated with inactivity [9,10], with risk of disease increased regardless of disability type [7]. Scoping reviews of literature report the use of mobile health (mHealth) to improve physical activity in special populations [11,12], with 1 randomized controlled trial showing significant increases in physical activity over time in internet-based physical activity interventions with adults with multiple sclerosis [13]. However, the research for people with physical disabilities is limited. Expansion of electronic health (eHealth) approaches to increase physical activity should continue to be investigated for all people with physical disabilities, with a focus on mobile apps as the market continues to expand.

There are 380,000 mobile apps related to health and fitness, with over 78,000 added in 2017 alone [14]. mHealth apps focusing on physical activity tend to lack theory or evidence-based content and messaging, and few of the commercially available apps have been actively studied for their effectiveness to increase physical activity or intent to continue activity [15,16]. However, there is evidence that eHealth interventions (internet-based and native mobile apps) that target physical activity resulted in increases in users' activity [17,18], and mobile physical activity app users are more likely to maintain behaviors than nonusers [19]. In addition, taking advantage of gamification features in mHealth apps resulted in improved user health and well-being [20].

Increased access to technology, specifically mobile apps, may be able to decrease prevalence of inactivity of users, including users with physical disabilities. Plow and Golding [21] found that the use of commercially available apps designed to promote self-management showed increases in planned physical activity in adults with musculoskeletal and neurological conditions. People with musculoskeletal conditions also adhered better to home therapy exercise routines and were more likely to increase function when prescribed programs were available on an app, compared with take-home paper instructions [22]. App use shows promise for people with physical disabilities, but apps are not generally designed specifically for these users, causing accessibility issues during use. Mobile devices' small screens with little negative space make using the device difficult for people with visual impairments or dexterity issues [23]. Therefore, people with physical disabilities have to adapt to the app rather than have an app that is intuitive for them. Active people with physical disabilities understand the unique physical activity needs of people with physical disabilities and have the experience to inform design of an app that is more likely to meet their needs at various levels of physical activity participation.

Expanding Landscape of Adapted Sport

In the United States, traditional recreational sport competitions have had to make room for nontraditional activities, such as functional fitness, obstacle course racing (OCR), and strongman competitions. Marathons and shorter races have seen registrations on a slow decline after peak registrations in 2013 [24]. Meanwhile, there is a growth in participation in less traditional OCR. Events such as the Spartan and Tough Mudder report 5 million and 2 million participants, respectively [24]. CrossFit, another nontraditional sport, has risen globally for the last 10 years. By the beginning of 2019, there were approximately 15,500 registered affiliates worldwide [25].

Parasport activities, sports such as wheelchair basketball and track events that are represented in Paralympic competition, are promoted through local parasports clubs, the Department of Veterans Affairs, and a variety of adapted sport-based nonprofit organizations. In recent years, people with physical disabilities have increased their recreational participation in nontraditional sports and are being recognized by organizers as consumers of nontraditional sport. In 2018, Spartan introduced the first ever Para Spartan Heats. Adaptive athletes (also known as people with physical disabilities) in the Elite Para division competed for a grand prize of US $10,000, and the event was covered by the Entertainment and Sports Programming Network, a major sports-broadcasting network. Although anecdotal, the expansion and visibility of people with physical disabilities in large nontraditional sport events is indicative of the growing interest of the community. As interest continues to grow, so might the consumer market for physical activity apps for them. To maximize the potential for proposed apps to support behavior change, specifically sustained physical activity participation, the apps should be designed to meet the needs and desires of

the target audience, the people with physical disabilities, and should use evidence-based design features.

User-Centered Design

From a commercial perspective, investment in design and development of an app is paramount because if a consumer is not engaged with the app, they will not return or recommend it to others. This also applies to mHealth—if the end user is not engaged, the intervention effectiveness is undermined [26]. Understanding the end-user requirements and the context of use increases a design's success rate of usability and, ultimately, the engagement necessary for an mHealth intervention implementation [27]. Currently, however, commercial apps do not integrate health intervention best practices, and mHealth interventions do not include the end user in the design process, reducing the potential effectiveness of the app to create sustained behavior change [26].

Focus groups can be a valuable tool in the early stages of the user-centered design (UCD) process to better understand what users want in the app. For a new app with broad potential reach, the type and number of actual users may be unclear until launch. Therefore, defining primary user characteristics is necessarily an iterative process [28]. This study recruited currently active people with physical disabilities who had used a fitness tracking app at least once to understand the unique needs of people with physical disabilities and to begin to build a taxonomy of features for reference in future focus groups of people with physical disabilities who might be less familiar with fitness tracking apps.

Objective

The indication that mobile app use increases physical activity adherence and the growth of participation of people with physical disabilities in nontraditional sport emphasizes the potential for mobile fitness tracking apps, designed specifically for users with physical disabilities, across a wide variety of physical activities to influence positive health outcomes. This study represents the first phase of UCD data collection for the development of an app for people with physical disabilities. The aim of this study was to explore user preferences for content, appearance, and operational features for a physical activity app for people with physical disabilities. Findings from this study will be used to inform the initial design phase of the mobile app.

Methods

Study Design

To encourage interaction and generation of ideas related to a proposed fitness tracking app for people with physical disabilities, focus groups were chosen. Focus groups offer an opportunity for discussion, debate, and idea sharing that has the potential to draw out more nuanced preferences of app content and features [29]. A total of 4 focus groups comprising adults with physical disabilities who were active in nontraditional sport (eg, no Paralympic equivalent) were conducted in the south, west, and mid-Atlantic regions of the United States throughout the months of March and April 2018. Focus groups were conducted by an experienced facilitator and an observer note taker. The facilitator guide was reviewed by a professional focus group moderator who made suggestions to improve wording and flow of questions. Focus group sizes averaged 4 participants. Institutional review board approval for data collection with active consent was obtained in advance through the University of Maryland.

Participants

Adults, aged 18 to 54 years (mean age 35.7, SD 9.2 years), who had a self-reported physical disability and were currently engaged in nontraditional sport activities were recruited through Crossroads Adaptive Athletic Alliance's network of fitness venues and via internet-based posting in closed Facebook groups built for adaptive athletes. A total of 15 people volunteered to participate. Participants were mostly white (n=13) and approximately half of all participants were female (n=9 female). Patients varied in mechanism of disability (n=8 acquired, n=6 congenital, n=1 both). The most prevalent nontraditional sport listed was CrossFit (n=15), and all participants engaged in at least one other sport. The demographic data, as collected on a questionnaire, for participants are presented in Table 1.

Table 1. Demographic characteristics of focus group participants across 4 focus group sessions and 3 regions in the United States.

Demographic characteristics	Focus groups participants (N=15), n (%)
Gender	
Female	9 (60)
Male	6 (40)
Age (years)	
18-29	3 (20)
30-39	7 (47)
40-49	4 (27)
50-59	1 (7)
Race or ethnicity	
White	13 (87)
African American	1 (7)
Asian	1 (7)
Region	
South	8 (53)
West	4 (27)
Mid-Atlantic	3 (20)
Disability	
Lower extremity amputation	4 (27)
Cerebral palsy	3 (20)
Spinal cord injury	3 (20)
Other (eg, club foot and limb paralysis)	3 (20)
Upper extremity amputation	1 (7)
Visual impairment	1 (7)
Mechanism	
Acquired	8 (53)
Congenital	6 (40)
Both	1 (7)
Current sport participation	
CrossFit	15 (100)
Parasport (ie, Paralympic equivalent available)	7 (47)
Olympic lifting	4 (27)
Baseball	1 (7)
Bodybuilding	1 (7)
Highland games	1 (7)
Obstacle course racing	1 (7)
Sailing	1 (7)
Strongman	1 (7)
Volleyball	1 (7)
Yoga	1 (7)

Procedures

Before the sessions began, participants were informed about the purpose of the groups, what they would be asked, the benefits and risks of participation, that participation was voluntary, and how their responses would be kept confidential. Participants signed and received a copy of the informed consent forms. Focus group sessions lasted approximately 2 hours and were audio recorded. Lunch was provided during the focus group sessions, but no other incentives were offered.

A semistructured guide was developed to frame the discussion. The session was designed to cover 4 broad topic areas. The first part of the session examined participants' introduction to adapted sport. The second part explored current app use. Participants were asked to describe the apps they used most frequently and what features they liked most. The question was used to elicit features that appealed to participants, regardless of app purpose. All participants were current app users with the most common apps mentioned being games or social media. However, apps organically introduced by participants also included fitness tracking apps. Regardless of app type, preferred features that were mentioned by the participants were included in the final ranking activity. In the third phase of the focus groups, participants were provided screenshots from 2 mobile apps: Gateway to Gold, designed by US Paralympics as a fitness tracking and recruiting tool specifically for people with physical disabilities, and Beyond the Whiteboard, the official app of CrossFit to track workouts and connect a geographically disparate fitness community. None of the participants had used either app, although a few mentioned that they had heard of Beyond the Whiteboard. On the basis of images alone, participants provided feedback and described functionality assumptions. In the final phase of the focus groups, participants were asked to consider their own experiences and the session's discussion to describe the features of an ideal fitness tracking app for people with physical disabilities. Some of the gamification and social interaction features participants described in this phase of the focus group were originally mentioned in the second phase of the focus group in which features of most frequently used apps were discussed. The end of each session included an activity in which app features mentioned by the participants throughout the session were displayed for participant review. Therefore, each focus group had a slightly different list of features to review. Table 2 lists these features such as offering broad disability selection options for user profiles. The participants independently rated priority features as *must have*, *nice to have*, or *not needed*. They were not required to vote on every feature.

Table 2. Focus group participants' identification and rankings of preferred features in a disability-specific workout app.

Theme and identified features[a]	"Must have"	"Nice to have"	"Not needed"	Total participants rating this feature
Profile				
Broad profile classifications of disability (vs more traditional narrow Paralympic classifications)	2	—	—	11
Customizable profile (avatar, location, social media handles, etc)	1	1	—	6
Subcategories of disability category available for selection on profile page	—	—	—	9
Workout-specific				
Detailed and strict movement standards for multiple disabilities for each posted workout	8	—	—	15
"Additional assistance needed"/modification options for workouts	8	—	—	15
Workout history (view by workout or calendar)	7	—	—	12
Workout search function (by primary muscle group/body part or specific movement)	2	3	—	8
Workout/training tips	2	—	—	6
Wholistic tracking (nutrition, sleep, etc)	—	2	—	13
Lift percentages precalculated based on 1RM[b], 3RM, or 5RM in the benchmark section	—	2	1	7
User experience				
Links to an in-app resources page (grants for equipment and travel, used adapted sport gear, link to Parasports club pages, etc)	5	—	—	13
Intuitive functions and navigation	4	—	—	13
Compatible with phone screen readers	2	—	—	4
User control of what others can see, privacy settings	2	—	—	9
Free	1	—	—	9
User video upload	1	—	4	9
Glossary of terms	—	3	—	6
Explanation on landing page of who the app is for/who is allowed to join	—	—	3	6
Gamification features				
Leaderboard sortable by location, disability, and gender	4	6	—	15
Goal setting in app	—	4	—	4
Share PRs[c] and workout comments in app	—	2	—	5
In-app prompts to move up a level or try a heavier weight (based on time since last attempt, comparison with others at the same level, or progress marked in other benchmarks)	—	2	—	13
Levels defined for each workout, with next level locked until previous one is completed	2	—	2	15
Profile badge for "level" achieved in app	—	3	2	9
In-app notifications (eg, PR, movement on leaderboard)	—	2	4	9
Option to make scores public or private	—	—	2	9
Social/user interaction				
Ability to network with and message other users	2	2	—	12
Share in-app successes on social media	—	3	—	15
Positive reinforcement from other users ("like" score, post emoji)	—	—	2	7
Compete with friends in app (eg, post own challenges)	—	—	2	7
Ability to create workout groups with other users	—	—	1	4

Theme and identified features[a]	"Must have"	"Nice to have"	"Not needed"	Total participants rating this feature
Ability to look up other athletes	—	—	1	10
Ability to comment on workout descriptions	—	—	1	6
Sync results to fitness tracker	—	—	1	6

[a]The numbers in each column reflect aggregated votes across all focus groups. Participants were asked to place colored stickers by features to represent *must have*, *nice to have*, and *not needed* app features but not required to vote on each feature. Some features received no stickers but remained on the table to indicate being mentioned during the focus group session.

[b]RM: repetition maximum, for example, 1RM is the maximum weight lifted by an individual for 1 repetition.

[c]PR: personal record.

Data Analysis

During the focus groups, notes taken by the facilitator and note taker were compared to ensure all mentioned features were included in the final activity. Participants were also asked to review the list to ensure nothing was missed or presented other than intended during conversations. Saturation was reached by the fourth focus group session. This list formed the basis of the results. In addition, focus group sessions were recorded and transcribed verbatim and verified for accuracy by a second listening. Using the processes inherent in thematic analysis [30], each focus group transcription was open coded by a single researcher (SHO) with the review of thematic analysis by coauthors (SLS and RSG). Codes related to the app features were then collated into potential themes and compared against coded data extracts. The full list of feature themes derived from the transcriptions was used to verify the completeness of the lists used during the focus group rating activity. No additional feature-based themes emerged from the data during this last step.

Results

Summary

Analysis of focus groups sessions revealed 5 overarching themes across the 34 recommended features and functions. The themes address preferences for (1) workout-specific features that were tailored to or searchable by disability, (2) user experience that was intuitive and accessible, (4) profile personalization options, (4) gamification features that allowed for competition with self and other users, and (5) social features that allowed increased interaction among users. Several of the user experience and social feature preferences mentioned by the participants were based on apps and app features that were not fitness related. The complete list and total number of votes of each feature aggregated across all focus groups can be found in Table 2.

Profile Features

Participants noted that the profile setup page for the comparison Gateway to Gold app used very specific Paralympic-based classification terms for the profile (eg, Les Autres and Dysmelia) that participants did not relate to or identify with. Given the fact that the proposed app would not be built specifically for Paralympic hopefuls, participants believed the categories should be less specific and more understandable to all users. They also wanted the ability to personalize their profile page.

Participants mentioned the preference for customizable profile pages including video upload:

> *…the humor part of it, it's part of expression, personalizing it. That's another thing just being able to personalize something. I don't know that's the girly side of me, if I can make it pink, would I? Or whatever, you know, something to personalize it. Everybody wants it to be their own somehow.* [Participant 2]

> *I know sometimes if you're posting videos of what you're doing it might inspire someone else to do that or you'll see it and say, oh I can do that, as opposed to reading it. It makes a different connection.* [Participant 10]

Workout-Specific Features

Both comparison app screenshots showed a library of available workouts. In a proposed app that was designed for a wide variety of people with physical disabilities, participants were primarily concerned with understanding how to do a new workout or fitness movement contained in the library. They wanted workouts with links to short videos contained in an in-app library that demonstrated a movement with multiple adaptation options, clearly described standards for proper completion of movements (eg, a squat for a below the knee amputee would mean hips drop below parallel on each repetition) differentiated by broad classifications of athletes, and provided recommendations for modifying the movements if the given standards were too hard for the user. It was also important for the participants to be able to see a history of their workouts to quickly determine improvement and set new goals. The example most frequently given was when a workout was opened from the search page, the dates and scores of all previous attempts would be listed for the user who was logged in.

Participants mentioned the preference for in-app demonstration videos of a wide variety of movements and standards for a wide variety of disability categories (Rx: how a movement or workout is prescribed by the coach or fitness instructor; denotes the weight, movements, and number of repetitions in a workout; used by participants generally referring to workouts similar to CrossFit):

> *…what I mean by that is there's a difference between him and her. Right? And between me and you. So, what my thought is you could click on your ability level basically and it would have a video or something*

> *like that of how you do it. So, it wouldn't just have the one. You know what I'm saying?* [Participant 6]
>
> *Or like a section with videos posted by the app, to show you how to do a movement, kind of like an instructional video from professionals instead.* [Participant 10]
>
> *The exact thing that the Open does specifically. They have the different variations that give a video of the exact workout for the RX athlete. Then you have specific individual photos to show the scaled versions for whoever it is, whatever category and whatever person it is.* [Participant 8]

User Experience

Participants spoke of the unique position many adaptive athletes find themselves in as the only person they know attempting sport and how that alone can be a barrier to participation. They recommended a resources page with links to adaptive sport event calendars, popular adaptive devices used in sport, and grant opportunities to enable participation. From a user experience perspective, participants noted the text-heavy nature of both Gateway to Gold and Beyond the Whiteboard apps. The amount of information on each screen was distracting, and they would prefer an app that is more visual and more intuitive. A participant complained:

> *It does have pretty small text. It's too busy. Too much information that I don't care.* [Participant 2]

A positive feature noted by a participant with visual impairments was the large text in the buttons and use of contrasting colors in the Gateway to Gold app that made navigation easier for her than in other apps in which she had to rely solely on her screen reader.

The participants explained why a resource page was important to adaptive athletes:

> *Let's just say in my condition right now, I want to get into wheelchair basketball and if I'm watching other people on this app doing these things. If I'm going, "I know my wheelchair isn't that one; it can't do that," again there's that resource. Click on this; its highlighted. This is where you get that type of resource.* [Participant 2]
>
> *I know as a blind person I never really heard much about sports or inclusive sports.* [Participant 11]

Gamification

Participants stated that in addition to tracking their own progress, they wanted the ability to compare their results with others, and more importantly, others similar to themselves. Several participants used FitBit, Wodify, and MyFitnessPal but were frustrated that wheelchair versus run options or 1-arm lifts were not available selections, making it impossible to compare themselves with others on the app. The most appealing option, therefore, was a leaderboard that would allow them to sort scores by disability type, adding competition with others *like them* to the app and making it unique from all other apps currently used. Participants wanted to be able to sort by disability and click on profiles of other athletes. Additional gamification features mentioned were related to scores in workouts such as personal record notifications that matched features they already used in other apps.

The participants expressed their opinions on the preferred use of an in-app leaderboard:

> *You compete against people who are, have around the same abilities as you, so it would be nice to see where you fall. I think to be able to toggle between a very specific category and then a really broad category.* [Participant 10]
>
> *It's the competitive piece to see how you stack up against others.* [Participant 7]

Social Features

Despite all participants having multiple social media accounts, participants did not prioritize the ability to connect to those accounts or enable cross-posting across apps. Features participants liked in other apps that they thought this app should integrate included the ability to post notes, graphics interchange format files, and images to other people's accounts or feeds within the app and the ability to interact with each other concerning the workouts. They liked encouragement that a simple emoji on a workout score provided and believed that finding someone with a similar disability doing the same workouts might increase their motivation or performance over time.

Discussion

Principal Findings

The results of this study support the need for a physical activity app designed for and with people with physical disabilities. Focus group participants discussed and ranked a wide variety of features and content for a proposed physical activity app for people with physical disabilities. Greatest consensus among participants was achieved for features that would make this app unique to people with physical disabilities or offered mechanisms for goal setting and competition, such as clear and detailed movement examples and standards for a wide variety of disabilities, disability-specific modification or progression options for movements that are currently too difficult for the user, resources page unique to adaptive sport, and a sortable leaderboard that includes a filter for disability category.

The features listed as most desirable by participants mirror those found to, empirically, support behavior change. Scoping studies reveal that apps do not intentionally incorporate health behavior constructs [31] nor do they score well on instruments designed to evaluate inclusion of behavior change theory strategies [32]. Because apps on the market do not effectively integrate constructs or strategies of behavior change, consumers may need to use multiple fitness apps to affect behavior change [33]. The inclusion of behavior change techniques, derived from multiple theories and incorporated into a taxonomy for use [34], has been shown to improve the likelihood of change and the maintenance of the newly acquired behavior [33,35,36]. These focus groups, although not explicitly discussing theoretically

based development, identified features and content most likely to effectively support behavior change.

The most common and effective behavior change techniques incorporated into commercially available physical activity apps were found in goal setting, behavior feedback, and behavior demonstration features [16,36-38]. Focus group participants listed the ability to comment on other users' workout results and workout videos, the ability to record and track progress, and the need for a wide variety of demonstration videos as priorities for their ideal fitness tracking app. They also commented frequently about the opportunity this app would give to be a part of a larger, connected adapted sport community and how that connectedness would offer a much wider set of examples for modifications to fitness movements based on disability, something that is not readily accessible to them now. Networkability—the degree to which the app supports a social networking function—has also shown increases in intention to continue using fitness tracking apps [39]. It is clear from the focus group discussions that people with physical disabilities believe a fitness tracking app designed for them would increase the sense of belonging, access to role modeling behavior either vicariously or through in-app interaction, and awareness of adaptive sport participation options.

Commercially available fitness apps incorporate gamification design principles to increase *relatedness* by integrating features that enable users to interact [40]. These same features—leaderboards and messaging and commenting on other users' posts or successes—are among the features focus group participants most desired. The application of gamification in health and fitness apps suggests that it can lead to a wide variety of positive behavior change impacts [20]. Gamification, specifically the use of leaderboards, has also been shown to significantly influence physical activity participation [41]. Developers marketing to people with physical disabilities should combine behavior change, technical, and design features to encourage increases in physical activity [37]. It is important to take into account unique desires of people with physical disabilities in the design process. Fitness tracking apps for people with physical disabilities should include demonstrations and clear instructions on how to perform suggested movements based on impairment. The user interface for people with physical disabilities, especially those with visual and dexterity impairments, should include high-contrast views with minimal text and large buttons for navigating the app.

Limitations

This study used a nonrandom purposive sample, collecting data through a network of fitness venues associated with a single nonprofit organization. This recruitment method may have introduced a selection bias; individuals who agreed to participate were already active in nontraditional sport, so their app needs might be different from a person with a physical disability just becoming active. One example of this was the extensive discussion of peer support social features during the focus groups sessions but the low rating of those features in the final activity. For active people with physical disabilities, peer modeling and virtual social support may be less important to sustained physical activity, especially if such support is provided through their sport participation. This may be enhanced by participation in traditional parasport team activities by half of the participants. All participants were active in CrossFit in addition to other sports. The average CrossFit athlete is of high socioeconomic status, whereas the average person with a physical disability is of lower socioeconomic status, which may have influenced the preference of certain features or content. Although the participant group was more homogenous with respect to race/ethnicity, the sample represented people with a wide range of disabilities, sporting interests, and an equitable number of males and females. Only 4 focus groups were conducted; however, data saturation was reached by the fourth focus group, and findings represented a diverse set of participant preferences [42]. Despite saturation of suggested features, the heterogeneity of participant needs was clear in the rating of app features. Only 2 features were considered *must have* by more than half of the participants.

Conclusions

UCD considers the needs and preferences of the end user at all stages of development. This is particularly important for anyone developing an app for people with physical disabilities; unless you have the lived experience of accessing apps as a person with physical disabilities, designing for this group is challenging. Besides providing input for a more usable platform, the process of using UCD for apps intended for people with physical disabilities empowers an often marginalized population to contribute to their own health. The results showed that features related to user experience, social engagement, and gamification are considered important to people with physical disabilities. Though none of the apps used as examples by the participants integrated theory, the features they highlighted as most desired were in line with the research-identifying attributes of quality apps that used behavior change techniques to influence positive physical activity behavior change. Such insights should inform the development of any fitness app designed to integrate users with disabilities as a primary user base. The heterogeneity of preferences when asked to rate each feature indicates that future research is needed to understand how disability type and physical activity participation affect app feature and content preference of the people with physical disabilities.

Acknowledgments

The authors would like to thank Crossroads Adaptive Athletic Alliance for the focus group travel funding and Project Momentum for providing trained focus group notetakers.

Conflicts of Interest

None declared.

References

1. Martin JJ, Whalen L. Self-concept and physical activity in athletes with physical disabilities. Disabil Health J 2012 Jul;5(3):197-200. [doi: 10.1016/j.dhjo.2012.03.006] [Medline: 22726861]
2. Sothern MS, Loftin M, Suskind RM, Udall JN, Blecker U. The health benefits of physical activity in children and adolescents: implications for chronic disease prevention. Eur J Pediatr 1999 Apr;158(4):271-274. [doi: 10.1007/s004310051070] [Medline: 10206121]
3. Warburton DE, Bredin SS. Health benefits of physical activity. Can Med Assoc J 2006 Sep 26;175(7):777 [FREE Full text] [doi: 10.1503/cmaj.1060147]
4. Wilhite B, Shank J. In praise of sport: promoting sport participation as a mechanism of health among persons with a disability. Disabil Health J 2009 Jul;2(3):116-127. [doi: 10.1016/j.dhjo.2009.01.002] [Medline: 21122750]
5. Everson-Hock ES, Green MA, Goyder EC, Copeland RJ, Till SH, Heller B, et al. Reducing the impact of physical inactivity: evidence to support the case for targeting people with chronic mental and physical conditions. J Public Health (Oxf) 2016 Jun;38(2):343-351. [doi: 10.1093/pubmed/fdv036] [Medline: 25812977]
6. Diament M. Disability Scoop. 2014. CDC: 1 In 2 With Disabilities Physically Inactive URL: https://www.disabilityscoop.com/2014/05/07/cdc-physically-inactive/19343/ [accessed 2015-09-01]
7. Centers for Disease Control and Prevention. 2014. Nutrition, Physical Activity, and Obesity: Data, Trends and Maps URL: https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html [accessed 2017-04-01]
8. Centers for Disease Control and Prevention. Physical Activity: Data and Statistics URL: https://www.cdc.gov/physicalactivity/data/facts.htm [accessed 2018-02-01]
9. Carroll DD, Courtney-Long EA, Stevens AC, Sloan ML, Lullo C, Visser SN, Centers for Disease Control and Prevention (CDC). Vital signs: disability and physical activity--United States, 2009-2012. MMWR Morb Mortal Wkly Rep 2014 May 9;63(18):407-413 [FREE Full text] [doi: 10.15585/mmwr.mm6429a2] [Medline: 24807240]
10. Ginis KA, Jörgensen S, Stapleton J. Exercise and sport for persons with spinal cord injury. PM R 2012 Nov;4(11):894-900. [doi: 10.1016/j.pmrj.2012.08.006] [Medline: 23174556]
11. Birkhoff S, Moriarty H. Interventions using smartphone health apps across various populations: an integrative review of the literature. J Inform Nurs 2016;1(1):13-24 [FREE Full text]
12. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of ehealth and mhealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
13. Motl RW, Dlugonski D, Wójcicki TR, McAuley E, Mohr DC. Internet intervention for increasing physical activity in persons with multiple sclerosis. Mult Scler 2011 Jan;17(1):116-128. [doi: 10.1177/1352458510383148] [Medline: 20921239]
14. Pohl M. Research2Guidance. 2017. 325,000 Mobile Health Apps Available in 2017 – Android Now the Leading mHealth Platform URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017 [accessed 2019-01-27] [WebCite Cache ID 75k91wFd6]
15. Ahmed OH, Lee H, Marchant H, Jones R, Hall EE. Navigating the new landscape of apps: overcoming the challenge of poor quality apps in sport and exercise medicine. Br J Sports Med 2018 Oct;52(20):1283-1284. [doi: 10.1136/bjsports-2017-097911] [Medline: 29146753]
16. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
17. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Mar 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
18. McIntosh JR, Jay S, Hadden N, Whittaker PJ. Do e-health interventions improve physical activity in young people: a systematic review. Public Health 2017 Jul;148:140-148. [doi: 10.1016/j.puhe.2017.04.001] [Medline: 28482322]
19. Wang Q, Egelandsdal B, Amdam GV, Almli VL, Oostindjer M. Diet and physical activity apps: perceived effectiveness by app users. JMIR Mhealth Uhealth 2016 Apr 7;4(2):e33 [FREE Full text] [doi: 10.2196/mhealth.5114] [Medline: 27056639]
20. Johnson D, Deterding S, Kuhn KA, 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]
21. Plow M, Golding M. Using mhealth technology in a self-management intervention to promote physical activity among adults with chronic disabling conditions: randomized controlled trial. JMIR Mhealth Uhealth 2017 Dec 1;5(12):e185 [FREE Full text] [doi: 10.2196/mhealth.6394] [Medline: 29196279]
22. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
23. Yu D, Parmanto B, Dicianno B. An mhealth app for users with dexterity impairments: accessibility study. JMIR Mhealth Uhealth 2019 Jan 8;7(1):e202 [FREE Full text] [doi: 10.2196/mhealth.9931] [Medline: 30622096]

24. Dern N. The Atlantic. 2018. Why Are Obstacle-Course Races So Popular? URL: https://www.theatlantic.com/health/archive/2018/07/why-are-obstacle-course-races-so-popular/565130/ [accessed 2019-01-27] [WebCite Cache ID 75kb3JPxR]
25. Henderson S. Morning Chalk Up. 2018. CrossFit's Explosive Affiliate Growth by the Numbers URL: https://morningchalkup.com/2018/10/23/crossfits-explosive-affilaite-growth-by-the-numbers/ [accessed 2019-01-28] [WebCite Cache ID 75kboF4J8]
26. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
27. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 2;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
28. Kujala S, Kauppinen M. Identifying and Selecting Users for User-Centered Design. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction. 2004 Presented at: NordiCHI'04; October 23-27, 2004; Tampere, Finland p. 297-303. [doi: 10.1145/1028014.1028060]
29. Gibbs A. Social Research Update - University of Surrey. 1997. Focus Groups URL: http://sru.soc.surrey.ac.uk/SRU19.html [accessed 2019-08-29]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. 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]
32. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
33. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
34. 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]
35. 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 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
36. Yang CH, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
37. Dallinga J, Janssen M, van der Werf J, Walravens R, Vos S, Deutekom M. Analysis of the features important for the effectiveness of physical activity-related apps for recreational sports: expert panel approach. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e143 [FREE Full text] [doi: 10.2196/mhealth.9459] [Medline: 29914863]
38. 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]
39. Lee HE, Cho J. What motivates users to continue using diet and fitness apps? Application of the uses and gratifications approach. Health Commun 2017 Dec;32(12):1445-1453. [doi: 10.1080/10410236.2016.1167998] [Medline: 27356103]
40. Jozani MM, Maasberg M, Ayaburi E. Slayers vs Slackers: An Examination of Users' Competitive Differences in Gamified IT Platforms Based on Hedonic Motivation System Model. In: Proceedings of the International Conference on Learning and Collaboration Technologies. 2018 Presented at: LCT'18; July 15-20, 2018; Las Vegas, NV, USA p. 164-172. [doi: 10.1007/978-3-319-91152-6_13]
41. Thorsteinsen K, Vittersø J, Svendsen GB. Increasing physical activity efficiently: an experimental pilot study of a website and mobile phone intervention. Int J Telemed Appl 2014;2014:746232 [FREE Full text] [doi: 10.1155/2014/746232] [Medline: 24963290]
42. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field Methods 2016 Jul 24;29(1):3-22. [doi: 10.1177/1525822X16639015]

Abbreviations

eHealth: electronic health
mHealth: mobile health
OCR: obstacle course race
RM: repetition maximum
PR: personal record
UCD: user-centered design

Edited by G Eysenbach; submitted 13.06.19; peer-reviewed by M Kittelson, A Silva, G Signorelli; comments to author 20.07.19; revised version received 05.08.19; accepted 19.08.19; published 11.10.19.

Please cite as:
Olsen SH, Saperstein SL, Gold RS
Content and Feature Preferences for a Physical Activity App for Adults With Physical Disabilities: Focus Group Study
JMIR Mhealth Uhealth 2019;7(10):e15019
URL: https://mhealth.jmir.org/2019/10/e15019
doi:10.2196/15019
PMID:31605518

Original Paper

Predicting Energy Expenditure During Gradient Walking With a Foot Monitoring Device: Model-Based Approach

Soon Ho Kim[1], BA; Jong Won Kim[2], PhD; Jung-Jun Park[3], PhD; Myung Jun Shin[4], MD, PhD; MooYoung Choi[1], PhD

[1]Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University, Seoul, Republic of Korea

[2]Department of Healthcare Information Technology, Inje University, Gimhae, Republic of Korea

[3]Division of Sport Science, Pusan National University, Busan, Republic of Korea

[4]Department of Rehabilitation Medicine, Pusan National University Hospital, Busan, Republic of Korea

Corresponding Author:
Jung-Jun Park, PhD
Division of Sport Science
Pusan National University
2, Busandaehak-ro 63beon-gil, Geumjeong-gu
Busan, 46241
Republic of Korea
Phone: 82 10 3285 1626
Email: jjparkpnu@pusan.ac.kr

Abstract

Background: Many recent commercial devices aim at providing a practical way to measure energy expenditure. However, those devices are limited in accuracy.

Objective: This study aimed to build a model of energy consumption during walking applicable to a range of sloped surfaces, used in conjunction with a simple, wearable device.

Methods: We constructed a model of energy consumption during gradient walking by using arguments based in mechanics. We built a foot monitoring system that used pressure sensors on the foot insoles. We did experiments in which participants walked on a treadmill wearing the foot monitoring system, and indirect calorimetry was used for validation. We found the parameters of the model by fitting to the data.

Results: When walking at 1.5 m/s, we found that the model predicted a calorie consumption rate of 5.54 kcal/min for a woman with average height and weight and 6.89 kcal/min for an average man. With the obtained parameters, the model predicted the data with a root-mean-square deviation of 0.96 kcal/min and median percent error of 12.4%.

Conclusions: Our model was found to be an accurate predictor of energy consumption when walking on a range of slopes. The model uses few variables; thus, it can be used in conjunction with a convenient wearable device.

(JMIR Mhealth Uhealth 2019;7(10):e12335) doi:10.2196/12335

KEYWORDS

energy expenditure; physical activity; human walking; wearable devices; mHealth

Introduction

Physical inactivity, despite its well-known health risks [1,2], continues to be a serious public health issue [3]. Recently, various wearable devices, including wristbands and mobile phones, have offered a way to track physical activity throughout the day. Such devices can be used in ambulatory conditions by individuals or in clinical settings to monitor patients' physical activity.

Many of these devices use an accelerometer-based method to predict energy expenditure [4-6]. However, these methods are limited in precision [7]. A basic, common assumption used is that the calorie consumption rate is proportional to the walking velocity. A GPS tracker can then be used to measure the walking distance and then compute the total energy consumption. However, this method is limited in accuracy and may not be feasible indoors.

Literature Review

The energetics of human locomotion has been closely studied for decades. Early studies focused on energy expenditure during walking [8-12] and running [13-16], and made comparisons with the energy expenditures of other animals [17]. Most relevantly, studies on walking energetics found a proportional relationship between energy expenditure and the square of the velocity. These early studies showed that reasonable accuracy can be attained with simple relations, despite the complexity of the act of walking. More recently, detailed models of walking dynamics have been presented that examine more closely the mechanics of walking [18-24]. These biomechanical models aim to explain human gait patterns via energy minimization. Also studied have been movements of the arm [25,26] and the head and trunk [27], as well as gait patterns in special groups of interest [28,29]. Such models have also been used in the field of robotics in developing walking robots [30].

Previous studies were primarily of academic interest, although inexpensive commercial devices have recently been made available for personal or clinical use. Such devices offer noninvasive ways to measure daily caloric consumption, and they have been assessed by numerous validation studies in the literature [31-36]. The most common types of commercially available products include the wrist-worn accelerometer and devices based on heart rate monitors. Although these devices are good predictors of the number of steps and heart rate, accurate prediction of energy expenditure is yet to be achieved [37]. These validation studies test for various settings; however, they usually lack a discussion of the model or algorithm used in their predictions.

This study proposes a model of walking energetics applicable to a range of slopes. The model is based on a simple equation and uses data from a wearable device. The method uses a foot monitoring system that can sense footsteps, which allows for direct measurement of step frequency. We found that a high-accuracy model can be developed for a range of upward and downward slopes. The fact that it is based on a direct measurement of footsteps allows the device to be versatile and applicable to diverse walking situations. The ability to track expenditure while walking on sloped surfaces is helpful for sloped outdoor ground and also indoor use of stairs or sloped treadmills.

Methods

Experimental Procedure

For model development and validation, an experiment was devised in which 73 healthy participants (34 female, 39 male) walked on a treadmill. The participants had a mean age of 43.6 (SD 15.0) years, mean height of 168.3 (SD 10.5) cm, and mean weight of 68.1 (SD 12.1) kg. Participants were selected from healthy volunteers (age 20 to 60 years) who registered in the department of Sport Science, Pusan National University, Busan, Korea. We excluded participants who had cardiovascular, musculoskeletal, or neurological disorders to avoid any confounding factors or biases. The participants were asked to walk on a treadmill at various values of the incline angle, *Theta*, and speed, *v*. Specifically, the angle was taken to be 0° (indicating no incline); 4°, 9°, and 14° (uphill); and −4°, −9°, and −14° (downhill). It was observed that calorie consumption took approximately 30 seconds to stabilize to a linear rate while walking. Each walking measurement lasted approximately 5 minutes to ensure a sufficiently long sample.

Calorie consumption was measured with a COSMED K4b2 portable gas analyzer system. This indirect calorimetry, based on the gas analyzer system, measures oxygen consumption, from which energy expenditure is computed. This method has been validated as an accurate measure through numerous comparative studies [38-40] and is used as a criterion measure in many validation studies [31-33,35-37]. The gas analyzer was worn during the treadmill experiment, and it recorded a time series of cumulative calorie consumption. To eliminate noise associated with the beginning and end of the experiment, we discarded data for the first 50 seconds and the final 10 seconds before computing the energy consumption rate. Then the basal metabolic rate [41] was subtracted to obtain the energy expenditure associated with walking, which is denoted by *P*.

Each participant also wore a foot monitoring system, consisting of shoe insoles equipped with eight pressure sensors. The insole used was a prototype developed by 3L Labs (Seoul, Korea), and provided to us for research purposes. A depiction of the foot monitoring system and the experimental setup is given in Figure 1. A Fitbit Surge, a wrist-worn accelerometer device, was also worn by each participant to compare the accuracy of its caloric consumption prediction. This study was approved by the Institutional Review Board of Pusan National University, Busan, Korea. All participants provided written informed consent (PNU IRB/2015_33_HR).

A value of 0, 1, or 2 indicated the pressure on each of the pressure sensors and was recorded with a frequency of 10 Hz, which resulted in an array of 16 integers for each time step of 0.1 s. A snippet from example data is shown in Figure 2. From the pressure sensor data, we were able to extract the step frequency, *f*. We performed this by examining the sum of the pressure sensor values at each time step. An example is shown in Figure 3. Although it is natural to consider the foot to be off the ground when this sum is 0, this can result in erroneous results if one or more of the pressure sensors remain at a value above 0 throughout the entire step cycle, either due to a faulty sensor or residual pressure. We found that better accuracy was achieved when high and low thresholds were used. This was done by first assigning the on-ground status to the first-time step, and then sequentially assigning either the on-ground or off-ground status to each following time step. If the previous time step was on-ground and the pressure sum was below the lower threshold, we assigned the off-ground status to that time step; if the threshold was not crossed, the time step was left in on-ground status. If the previous status was off-ground, the on-ground status was assigned if the pressure sum was above the upper threshold, and the off-ground status was assigned otherwise. Threshold values between 1 and 10 were tested and compared with manually assigned steps. Lower and upper threshold values of 2 and 5, shown in Figure 3, were found to produce accurate results.

Figure 1. Illustration of the foot monitoring system (left) and a picture of a participant walking on an uphill treadmill wearing the K4b2 portable gas analyzer (right).

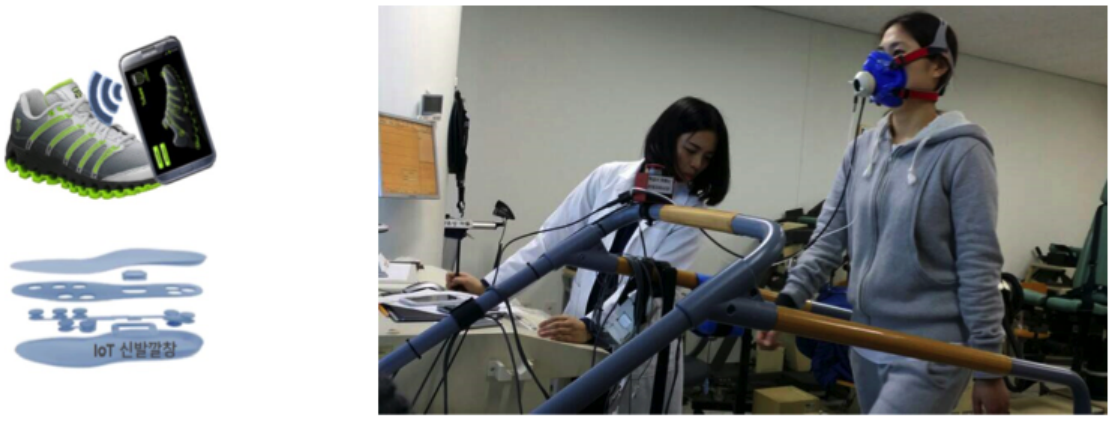


Figure 2. A sample of 4 seconds of raw data from the pressure sensors. The vertical position of each number of the array indicates the time, ordered from top to bottom at an increment of 0.1 seconds. Each column denotes a sensor, with left foot and right foot separated. The colored portions indicate when our algorithm decided the foot was off the ground.

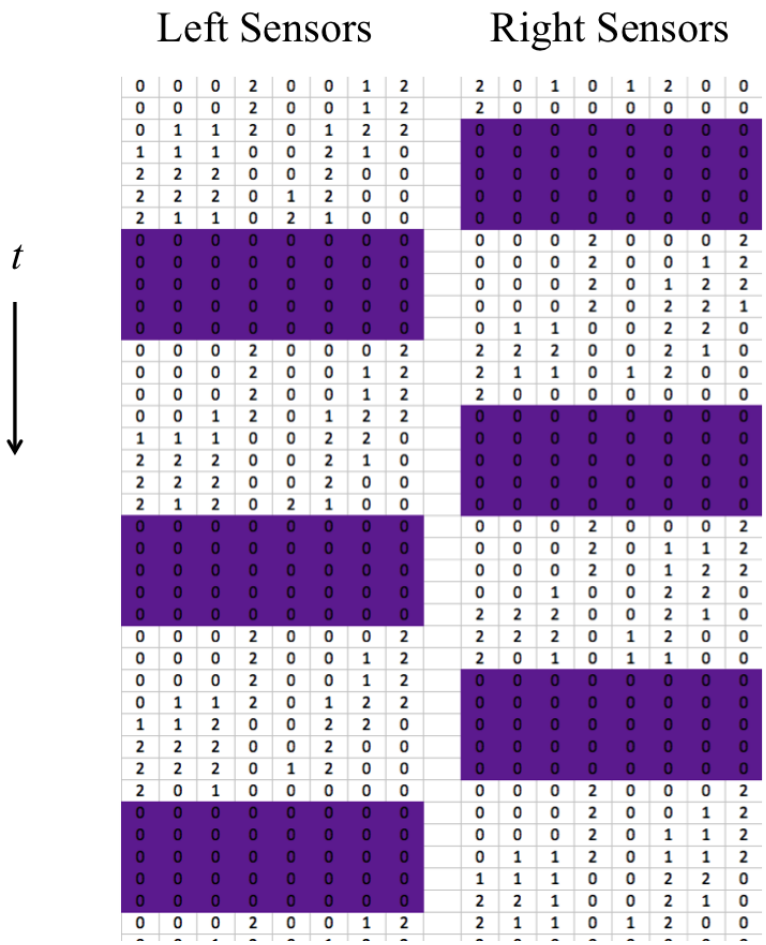
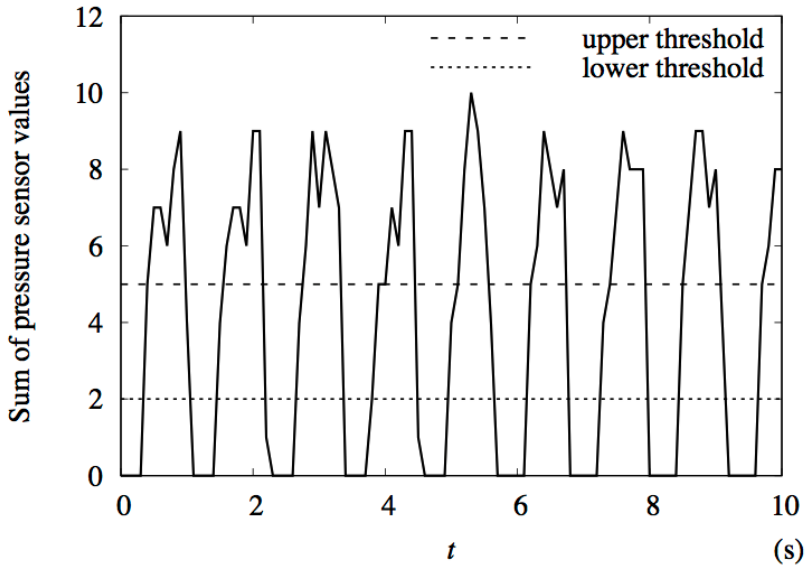


Figure 3. Graph of total pressure from the left foot sole over an interval of 10 seconds obtained from the foot monitoring system from the same data as presented in Figure 2. The two dashed lines indicate the upper and lower thresholds used to calculate the step frequency.



After assigning a status to each time step, we counted the number of transitions from the on-ground to off-ground status and divided it by the time interval to obtain the frequency. As with the gas analyzer data, we omitted data for the first 50 s and the final 10 s. Only one shoe insole is required to calculate the step frequency; however, we used the average of both sides in this study.

Model

Our model was constructed from considering the energy changes involved in walking. Suppose a participant with body mass *Mu* is walking with average speed *Nu* on a surface inclined by *Theta* from the horizontal. The participant is swinging their legs with frequency *f*. The energy consumption rate, *Rho*, is given by equation 1 (Figure 4). Here positive and negative values of the slope, *Theta*, of the walking surface correspond to walking uphill and downhill, respectively. Rho_K and Rho_U are rates of changes in the kinetic energy and in the potential energy, respectively, whereas coefficients $Gamma$,b_{Tau}, b_1, and Rho_0 are parameters to be determined empirically from the data. The energy change rates for Rho_K and Rho_U are given in equation 2 (Figure 4). In the following, we give an explanation of each term in consideration of energy.

Kinetic Energy Component

We first consider walking on a horizontal surface (ie, *Theta*=0). When walking on a treadmill, the upper body moves in a relatively constant velocity, with the moving legs supporting this movement. The legs swing back and forth relative to the upper body's position, undergoing an acceleration-deceleration cycle. We postulated that the energy expenditure was proportional to the kinetic energy change of the legs. The work done on the legs during each walking cycle is given by equation 3 (Figure 4). Here *m* is the mass of each leg, v_0 is the maximum speed of each leg's center of mass, and the factor of 4 accounts for the two legs each undergoing acceleration and then deceleration. This differs from the assumption that the legs swing like a pendulum, in which case gravity would do the work.

Since we usually have no way to easily measure leg mass or leg velocity, we defined two ratios: (1) the ratio α of the leg mass, *m*, to the body mass, *M;* and (2) the ratio β of the maximum velocity, v_0, of the leg to the average walking speed, *v* (equation 4 in Figure 4).

This allowed us to rewrite equation 3 as , giving an expression for the work done per cycle. Assuming that the human body converts chemical energy into kinetic energy with efficiency η_K, the energy consumption rate due to the kinetic energy is given by equation 5 (Figure 4). In writing the right-hand side of equation 5, the measurable terms are grouped into P_K as in equation 2, whereas the rest are grouped into dimensionless coefficient γ, given by equation 6.

Potential Energy Component

When walking on a horizontal surface perpendicular to the direction of gravity, there is no net change in potential energy. It changes when the subject is walking up or down a slope. We first considered upward inclines. When one walks up a slope of angle *Theta* at speed *v* parallel to the surface, their potential energy, *U*, changes at a rate $dU/dt=P_U$, given by equation 2 (Figure 4). For simplicity, we further assumed that when walking up a slope, additional energy proportional to this term is required. Accordingly, the energy expenditure rate associated with the changing potential energy is given by b_0P_U, where b_0 is the inverse of the efficiency, η_U, (equation 7 in Figure 4) with which the body converts stored energy to potential energy.

Figure 4. List of equations of the model of energy expenditure during walking.

$$P = \begin{cases} \gamma P_K + b_0 P_U + P_0 & \text{if } \theta \geq 0 \\ \gamma P_K + b_0 P_U + b_1 P_0^{-1} P_U^2 + P_0 & \text{if } \theta < 0 \end{cases} \quad (1)$$

$$P_K = 2Mv^2 f, \quad P_U = Mgv \sin\theta \quad (2)$$

$$W = 4 \times \frac{1}{2} m v_0^2 \quad (3)$$

$$\alpha \equiv \frac{m}{M}, \quad \beta \equiv \frac{v_0}{v} \quad (4)$$

$$\frac{1}{\eta_K} W f = 2\gamma M v^2 f = \gamma P_K \quad (5)$$

$$\gamma \equiv \frac{\alpha \beta^2}{\eta_K} \quad (6)$$

$$b_0 = \frac{1}{\eta_U} \quad (7)$$

$$b_0 P_U + b_1 P_0^{-1} P_U^2 \quad (8)$$

$$|P - P'|/P' \quad (9)$$

$$v \sim fh \quad (10)$$

One might consider simply using the same formula for downhill inclines, in which case the term b_0P_U=b_0Mgv sin*Theta* becomes negative. This would imply that when walking downslope, the change in potential energy can be converted into kinetic energy, thereby subtracting from the total energy cost. However, this leads to a nonsensical result for higher slopes, as it can lead to negative energy consumption. When a downhill slope is steeper than a certain angle, the subject would need to exert a frictional force to prevent from falling forward or walking too fast. Therefore, b_0P_U does not provide an adequate description of the energy expenditure in this case.

Figure 5 and 6 present scatterplots of the data in the three-dimensional space (P_K,P_U,P) for women and men, respectively. This visualization shows that P first decreases then increases as P_U is decreased from zero. Such a parabolic shape indicates the presence of a quadratic term; thus, we added to P a term proportional to P_U^2. The energy expenditure associated with potential energy in the case of downhill walking is given by equation 8 (Figure 4). The second term is multiplied by P_0^{-1} so that the coefficient b_1 is kept dimensionless. In other words, b_1 is the coefficient of the quadratic term in the case of downhill walking in units of P_0. This leads to the full model, described by equation 1 (Figure 4).

Figure 5. Three-dimensional scatterplot of data (dots) and model prediction (lines) of P versus P_U and P_K for women.

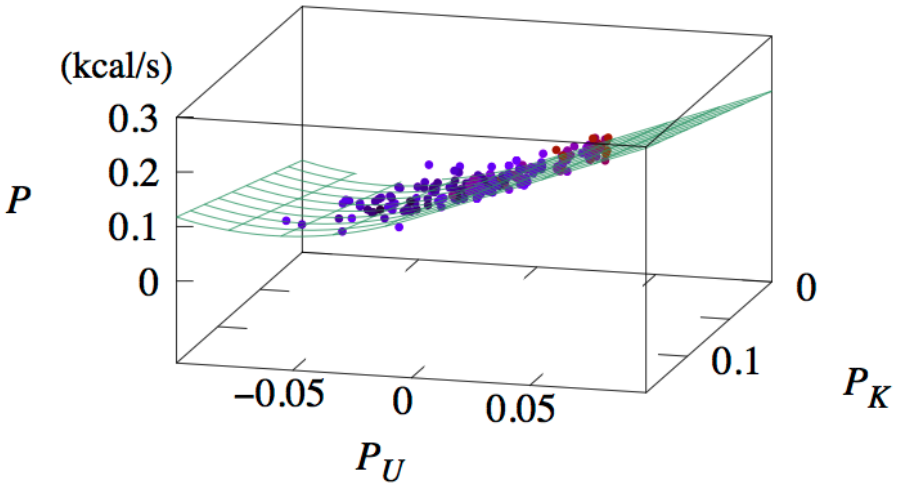
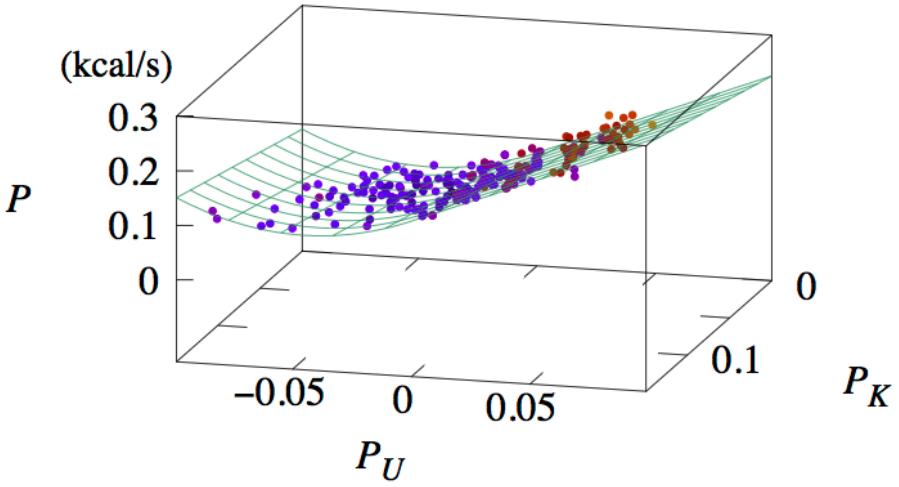


Figure 6. Three-dimensional scatterplot of data (dots) and model prediction (lines) of P versus PU and PK for men.



Linear Regression

The preceding model described leaves parameters Υ,b_0, b_1, and P_0 to be determined. We obtained these parameters by first taking data for flat and uphill surfaces (*Theta*≥0) and performing multiple linear regression through the use of the first equation in equation 1 (Figure 4) with Υ,b_0, and P_0 as fitting parameters. The adjusted R^2 value for the fits of both women and men was .83. Then b_1 was obtained via fitting the second equation of equation 1 (Figure 4) to flat and downslope data (*Theta*≤0). During this secondary fit, Υ,b_0, and P_0 were set constant at the values obtained earlier.

Results

The full set of coefficients, obtained through linear regression, is given in Table 1. The dependency of P on P_K and P_U is represented by the surfaces in Figures 5 and 6 . Due to the piecewise functional form of the model (equation 1 in Figure 4), the prediction plane has no curvature for P_U>0 but does in the region P_U<0.

Table 1. Coefficients for the full model reported with the root-mean-square deviation (RMSD) on comparison with data. The values were obtained by two linear regressions.

Coefficient	Units	Women	Men
γ	—	0.662	0.517
b_0	—	1.591	1.694
b_1	—	0.575	1.086
P_0	kcal/s	0.042	0.058
RMSD	kcal/s (kcal/min)	0.016 (0.96)	0.016 (0.96)

The fit resulted in a root-mean-square deviation (RMSD) of 0.96 kcal/min for both women and men. A boxplot of the percentage errors of all trials is given in Figure 7, in which the errors have been calculated according to equation 9 in Figure 4.

Here P is the prediction by the method whereas P' is the standard given by the gas analyzer. The median errors were 16.9% for women, 11.2% for men, and 12.4% for both groups. These errors are substantially lower than those found in a validation study for multiple commercial devices, which yielded median accuracies of 28.6% to 35.0% across devices for walking [37].

The predictions made by Fitbit Surge had an RMSD of 2.58 kcal/min (2.7 times that of the model) and a median percent error of 37.3% (3 times that of the model). However, this high error was mostly due to inaccuracies in sloped walking. When restricted to flat surfaces, the Fitbit Surge's accuracy increased dramatically, whereas the model's accuracy increased moderately. The Fitbit Surge's RMSD on flat surfaces was 1.82 kcal/min (2.3 times that of the model, 0.79 kcal/min), and the median percent error was 18.4% (1.6 times that of the model, 11.2%). Distributions of percent errors are portrayed with boxplots in Figure 7.

Before discussing the implications of these results, we note that the variables v and f are not independent. If l is the average length of a step, then $v = f\,l$. Assuming the approximate relation $h \approx l$, where h is the subject's height, we obtain $v \sim fh$ (equation 10 in Figure 4). This relation was observed in the data, as shown in Figure 8.

Equation 7 implies that η_U=0.547 for women and 0.596 for men. In principle, Υ depends on α,β, and η_K. We assumed the average value of α=0.185 for women and 0.165 for men, obtained from an anatomical reference [42], and that $\eta_K=\eta_U$. Taking these values and the fitting result for Υ, we obtained from equation 6 (Figure 4) the ratio β with values 1.47 for women and 1.36 for men. This difference in the average may reflect the difference in the average height between women and men. Specifically, equations 4 and 10 (Figure 4) imply $\beta=v_0/v \sim v_0/f\,h$. The ratio of the value of β for women to that for men equaled 1.08, whereas the ratio of the average height of men to that of women equaled 1.11.

Figure 7. Boxplots of the percent errors of predictions made by the model and Fitbit Surge. Errors have been estimated via equation 9 in Figure 4.

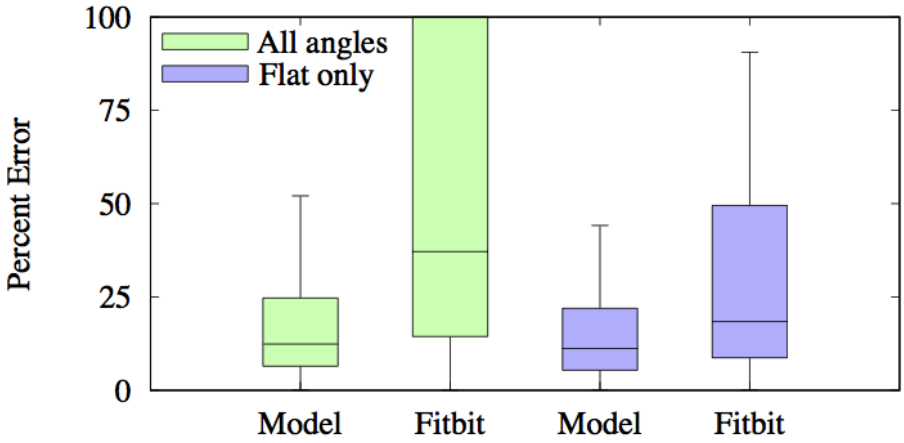
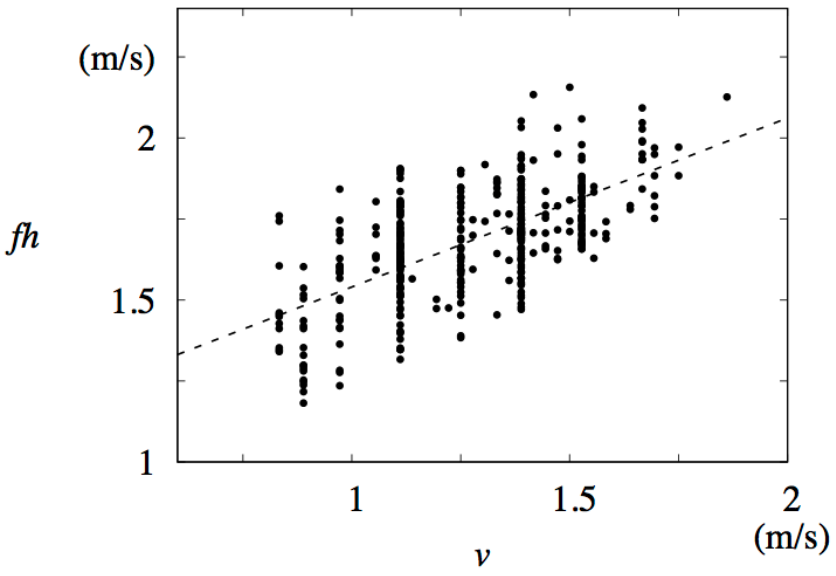


Figure 8. Step frequency, f, multiplied by height, h, plotted against average walking speed, v. Least squares fit line fh = 0.52v+1.02 (m/s) is also shown.



Discussion

Principal Results

We developed a model based on rates of change in kinetic and potential energies. In general, it predicts linear dependence of the energy consumption on these rates; in particular, it predicts quadratic dependence of the energy consumption on the potential energy change in the case of downhill walking. The method, used in conjunction with a foot monitoring system, predicts energy expenditure with an RMSD of 0.98 kcal/min and a median percent error of 12.4%, lower than those of wrist-worn commercial devices in predicting energy expenditure for walking. With one simple piecewise function, the model adequately predicts energy expenditure for walking in a wide range of the gradient.

Notice the differences in parameter values between women and men. The appreciable difference in the value of b_1 between men and women may result from the difference in walking posture; this is beyond the scope of this work and left for future study. In principle, the parameters are fit for each individual and should vary by subject. Thus, Table 1 presents average values of the coefficients within each gender. Even so, it is remarkable that a high degree of accuracy is observed.

Limitations

Although the model accounts for varying body mass and step frequency (cadence), this does not account for additional individual variations in parameter values due to walking gait and body dimensions. There may be ways to account for such variations without complicating the model. In addition, because the treadmill incline lies between 14° uphill and 14° downhill, we are not able to validate the model for more extreme slopes [43]. In addition, the method has not been tested and calibrated for outdoor walking or variable temperatures and altitudes. However, we believe that our pilot study provides a groundwork for follow-up studies under more ambulatory conditions.

Comparison With Prior Work

Prior studies have noted the strong correlations between P and v^2 for level walking [10]. The authors have also similarly considered additional energy expenditure when walking uphill, attributing it to vertical lift work. In contrast, our study proposes a simple formula that predicts energy consumption reasonably well for horizontal, uphill, and downhill surfaces within a unified framework. In addition, Cotes and Meade [9] made use of individual measurements, including resting metabolic rate and leg length. Our model shows that high accuracy can be achieved via reasonable assumptions used in conjunction with a wearable, mobile device.

Other existing studies have studied energy expenditure during uphill and downhill walking [43,44]. The authors reported a minimum energy cost when walking 10° downhill, which is consistent with our results. These studies did not incorporate varying walking speed and body weight, and relied on regression analysis with those variables kept constant. Our study offers a simple formula that applies to various walking speeds and subjects, while also accounting for the surface gradient.

Our method fits separately for women and men. Prior validation studies have found differences in the accuracy of devices between the two genders. A comparative validation study found that gender was one of the strongest predictors for accuracy, with a rate significantly higher for men than for women [37]. Our results suggest that similar error rates for both genders can be achieved.

Conclusions

We have developed a model that predicts energy expenditure during walking on a gradient surface between 14° uphill and 14° downhill, with an RMSD of 0.98 kcal/min. The model has been used in conjunction with a wearable device, the foot monitoring system, which directly measures footsteps. Thus, it offers an accessible method of measuring energy expenditure in realistic walking settings, where gradient walking is common. Future work may test equation 1 (Figure 4) in a wider range of values in the P_K–P_U space. Testing the method on outdoor walking is also desirable for further validation. Although not yet explored, the device could also be used in conjunction with other activity monitoring devices, such as wrist-worn ones, to produce more accurate measures of energy expenditure.

Acknowledgments

MY Choi acknowledges support from National Research Foundation of Korea through the Basic Science Research Program (grant No. 2016R1D1A1A09917318 and 2019 R1F1A1046285). MJ Shin and J-J Park acknowledge support from the Wearable Device R&D Project, Pusan National University Hospital, and KT (grant CMITKT-05).

Conflicts of Interest

None declared.

References

1. Nocon M, Hiemann T, Müller-Riemenschneider F, Thalau F, Roll S, Willich SN. Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis. Eur J Cardiovasc Prev Rehabil 2008 Jun;15(3):239-246. [doi: 10.1097/HJR.0b013e3282f55e09] [Medline: 18525377]

2. Wen CP, Wai JP, Tsai MK, Yang YC, Cheng TY, Lee M, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet 2011 Oct 1;378(9798):1244-1253. [doi: 10.1016/S0140-6736(11)60749-6] [Medline: 21846575]
3. Trost SG, Blair SN, Khan KM. Physical inactivity remains the greatest public health problem of the 21st century: evidence, improved methods and solutions using the '7 investments that work' as a framework. Br J Sports Med 2014 Jan 10;48(3):169-170. [doi: 10.1136/bjsports-2013-093372]
4. Maddison R, Gemming L, Monedero J, Bolger L, Belton S, Issartel J, et al. Quantifying human movement using the Movn smartphone app: validation and field study. JMIR Mhealth Uhealth 2017 Aug 17;5(8):e122. [doi: 10.2196/mhealth.7167] [Medline: 28818819]
5. Montoye HJ, Washburn R, Servais S, Ertl A, Webster JG, Nagle FJ. Estimation of energy expenditure by a portable accelerometer. Med Sci Sports Exerc 1983;15(5):403-407. [Medline: 6645869]
6. Meijer GA, Westerterp KR, Koper H, ten Hoor F. Assessment of energy expenditure by recording heart rate and body acceleration. Med Sci Sports Exerc 1989 Jun;21(3):343-347. [Medline: 2733585]
7. Chen KY, Sun M. Improving energy expenditure estimation by using a triaxial accelerometer. J Appl Physiol (1985) 1997 Dec;83(6):2112-2122 [FREE Full text] [doi: 10.1152/jappl.1997.83.6.2112] [Medline: 9390989]
8. Passmore R, Durnin JV. Human energy expenditure. Physiol Rev 1955 Oct 01;35(4):801-840. [doi: 10.1152/physrev.1955.35.4.801]
9. Cotes JE, Meade F. The energy expenditure and mechanical energy demand in walking. Ergonomics 1960 Apr;3(2):97-119. [doi: 10.1080/00140136008930473]
10. Dill DB. Oxygen used in horizontal and grade walking and running on the treadmill. J Appl Physiol 1965 Jan;20(1):19-22. [doi: 10.1152/jappl.1965.20.1.19] [Medline: 14257553]
11. Menier DR, Pugh LG. The relation of oxygen intake and velocity of walking and running, in competition walkers. J Physiol 1968 Aug;197(3):717-721. [doi: 10.1113/jphysiol.1968.sp008584] [Medline: 5666183]
12. van der Walt WH, Wyndham CH. An equation for prediction of energy expenditure of walking and running. J Appl Physiol 1973 May;34(5):559-563. [doi: 10.1152/jappl.1973.34.5.559]
13. Holt KG, Hamill J, Andres RO. Predicting the minimal energy costs of human walking. Med Sci Sports Exerc 1991 Apr;23(4):491-498. [Medline: 1905381]
14. Margaria R, Cerretelli P, Aghemo P, Sassi G. Energy cost of running. J Appl Physiol 1963 Mar 01;18(2):367-370. [doi: 10.1152/jappl.1963.18.2.367]
15. Zarrugh MY, Radcliffe CW. Predicting metabolic cost of level walking. Europ J Appl Physiol 1978;38(3):215-223. [doi: 10.1007/bf00430080]
16. Cavanagh PR, Williams KR. The effect of stride length variation on oxygen uptake during distance running. Med Sci Sports Exerc 1982;14(1):30-35. [doi: 10.1249/00005768-198201000-00006] [Medline: 7070254]
17. Steudel K. Limb morphology, bipedal gait, and the energetics of hominid locomotion. Am J Phys Anthropol 1996;99(2):345-355. [doi: 10.1002/(sici)1096-8644(199602)99:2<345::aid-ajpa9>3.3.co;2-m]
18. Minetti A, Alexander R. A theory of metabolic costs for bipedal gaits. J Theor Biol 1997 Jun;186(4):467-476. [doi: 10.1006/jtbi.1997.0407]
19. Alexander RM. A model of bipedal locomotion on compliant legs. Philos Trans R Soc Lond B Biol Sci 1992 Oct 29;338(1284):189-198. [doi: 10.1098/rstb.1992.0138] [Medline: 1360684]
20. Doke J, Donelan JM, Kuo AD. Mechanics and energetics of swinging the human leg. J Exp Biol 2005 Feb;208(Pt 3):439-445 [FREE Full text] [doi: 10.1242/jeb.01408] [Medline: 15671332]
21. Kuo A. Energetics of actively powered locomotion using the simplest walking model. J Biomech Eng 2002;124(1):113-120. [doi: 10.1115/1.1427703]
22. Kuo AD, Donelan JM, Ruina A. Energetic consequences of walking like an inverted pendulum: step-to-step transitions. Exerc Sport Sci Rev 2005 Apr;33(2):88-97. [Medline: 15821430]
23. Ruina A, Bertram JE, Srinivasan M. A collisional model of the energetic cost of support work qualitatively explains leg sequencing in walking and galloping, pseudo-elastic leg behavior in running and the walk-to-run transition. J Theor Biol 2005 Nov;237(2):170-192. [doi: 10.1016/j.jtbi.2005.04.004]
24. Osaki Y, Kunin M, Cohen B, Raphan T. Three-dimensional kinematics and dynamics of the foot during walking: a model of central control mechanisms. Exp Brain Res 2006 Aug 18;176(3):476-496. [doi: 10.1007/s00221-006-0633-1]
25. Donker SF, Mulder T, Nienhuis B, Duysens J. Adaptations in arm movements for added mass to wrist or ankle during walking. Exp Brain Res 2002 Jun 21;146(1):26-31. [doi: 10.1007/s00221-002-1145-2]
26. Zehr EP, Haridas C. Modulation of cutaneous reflexes in arm muscles during walking: further evidence of similar control mechanisms for rhythmic human arm and leg movements. Exp Brain Res 2003 Feb 6;149(2):260-266. [doi: 10.1007/s00221-003-1377-9]
27. Hirasaki E, Moore ST, Raphan T, Cohen B. Effects of walking velocity on vertical head and body movements during locomotion. Exp Brain Res 1999 Jul;127(2):117-130. [doi: 10.1007/s002210050781]
28. Hausdorff JM, Yogev G, Springer S, Simon ES, Giladi N. Walking is more like catching than tapping: gait in the elderly as a complex cognitive task. Exp Brain Res 2005 Apr 28;164(4):541-548. [doi: 10.1007/s00221-005-2280-3]

29. Blin O, Ferrandez A, Serratrice G. Quantitative analysis of gait in Parkinson patients: increased variability of stride length. J Neurol Sci 1990 Aug;98(1):91-97. [doi: 10.1016/0022-510x(90)90184-o] [Medline: 2230833]
30. Collins SH, Wisse M, Ruina A. A Three-Dimensional Passive-Dynamic Walking Robot with Two Legs and Knees. Int J Robotics Res 2016 Jul 02;20(7):607-615. [doi: 10.1177/02783640122067561]
31. 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 Tech 2013 Sep 05;37(7):456-462. [doi: 10.3109/03091902.2013.831135]
32. Brugniaux JV, Niva A, Pulkkinen I, Laukkanen RM, Richalet J, Pichon AP. Polar Activity Watch 200: a new device to accurately assess energy expenditure. Br J Sports Med 2008 Apr 15;44(4):245-249. [doi: 10.1136/bjsm.2007.045575]
33. Härtel S, Gnam J, Löffler S, Bös K. Estimation of energy expenditure using accelerometers and activity-based energy models—validation of a new device. Eur Rev Aging Phys Act 2010 Dec 2;8(2):109-114. [doi: 10.1007/s11556-010-0074-5]
34. 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]
35. 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 Cariol 2015 Apr;185:138-140. [doi: 10.1016/j.ijcard.2015.03.038]
36. Tucker WJ, Bhammar DM, Sawyer BJ, Buman MP, Gaesser GA. Validity and reliability of Nike + Fuelband for estimating physical activity energy expenditure. BMC Sports Sci Med Rehabil 2015 Jun 30;7(1):14. [doi: 10.1186/s13102-015-0008-7]
37. Shcherbina A, Mattsson C, Waggott D, Salisbury H, Christle J, Hastie T, et al. Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. J Pers Med 2017 May 24;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
38. Ainslie PN, Reilly T, Westerterp KR. Estimating human energy expenditure. Sports Med 2003;33(9):683-698. [doi: 10.2165/00007256-200333090-00004]
39. Levine JA. Measurement of energy expenditure. Public Health Nutr 2007 Jan 02;8(7a):1123-1132. [doi: 10.1079/phn2005800]
40. Haugen HA, Chan L, Li F. Indirect calorimetry: a practical guide for clinicians. Nutr Clin Pract 2017 Jan 06;22(4):377-388. [doi: 10.1177/0115426507022004377]
41. Mifflin MD, St JS, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. Am J Clin Nutr 1990 Feb 01;51(2):241-247. [doi: 10.1093/ajcn/51.2.241]
42. Plagenhoef S, Evans FG, Abdelnour T. Anatomical data for analyzing human motion. Res Q Exerc Sport 1983 Jun;54(2):169-178. [doi: 10.1080/02701367.1983.10605290]
43. Minetti AE, Moia C, Roi GS, Susta D, Ferretti G. Energy cost of walking and running at extreme uphill and downhill slopes. J Appl Physiol 2002 Sep 01;93(3):1039-1046. [doi: 10.1152/japplphysiol.01177.2001]
44. Minetti AE, Ardigo LP, Saibene F. Mechanical determinants of gradient walking energetics in man. J Physiol 1993 Dec 01;471(1):725-735. [doi: 10.1113/jphysiol.1993.sp019969]

Abbreviations

RMSD: root-mean-square deviation

Edited by G Eysenbach; submitted 27.09.18; peer-reviewed by M Stuckey, T Ebara, B Chaudhry, J Seitz; comments to author 01.04.19; revised version received 10.05.19; accepted 19.07.19; published 23.10.19.

Please cite as:
Kim SH, Kim JW, Park JJ, Shin MJ, Choi M
Predicting Energy Expenditure During Gradient Walking With a Foot Monitoring Device: Model-Based Approach
JMIR Mhealth Uhealth 2019;7(10):e12335
URL: https://mhealth.jmir.org/2019/10/e12335
doi:10.2196/12335
PMID:31647467

Original Paper

Mobile Health Coaching on Nutrition and Lifestyle Behaviors for Subfertile Couples Using the Smarter Pregnancy Program: Model-Based Cost-Effectiveness Analysis

Elsje C Oostingh[1], MD; Robbin H Ophuis[1], MSc; Maria PH Koster[1], MD, PhD; Suzanne Polinder[1], PhD; Hester F Lingsma[1], PhD; Joop SE Laven[1], MD, PhD; Régine PM Steegers-Theunissen[1], MD, PhD

Erasmus University Medical Center, Rotterdam, Netherlands

Corresponding Author:
Régine PM Steegers-Theunissen, MD, PhD
Erasmus University Medical Center
Postbus 2040
Rotterdam, 3000 CA
Netherlands
Phone: 31 107038255
Email: r.steegers@erasmusmc.nl

Abstract

Background: The health care costs for reproductive care have substantially increased with the use of in vitro fertilization (IVF) treatment. The mobile health (mHealth) coaching program Smarter Pregnancy is an effective intervention to improve nutrition and lifestyle behaviors and pregnancy rates in (sub)fertile couples, including those who undergo IVF treatment. Therefore, we hypothesize that this mHealth program can also reduce health care costs associated with IVF treatment.

Objective: This study aimed to evaluate the cost-effectiveness of the mHealth coaching program Smarter Pregnancy and compare it to usual care in women of subfertile couples who start their first IVF cycle.

Methods: This model-based cost-effectiveness analysis was performed on data from couples undergoing IVF treatment at the Erasmus MC, University Medical Center Rotterdam. A decision tree model was used to assess the incremental cost-effectiveness ratio (ICER) of ongoing pregnancies and costs of use of the mHealth program as compared to usual care. A probabilistic sensitivity analysis was performed to consider the uncertainty surrounding the point estimates of the input parameters.

Results: Based on our model including 793 subfertile women undergoing IVF treatment, use of the mHealth program resulted in 86 additional pregnancies and saved €270,000 compared to usual care after two IVF cycles, with an ICER of –€3050 (95% CI –3960 to –540) per additional pregnancy. The largest cost saving was caused by the avoided IVF treatment costs. Sensitivity analyses showed that the mHealth program needs to increase the ongoing pregnancy rate by at least 51% after two IVF cycles for cost saving.

Conclusions: The mHealth coaching program Smarter Pregnancy is potentially cost saving for subfertile couples preceding their first IVF treatment. Implementation of this mHealth program in routine preconception care for subfertile couples should be seriously considered, given the relatively low costs and promising cost-effectiveness estimates.

(JMIR Mhealth Uhealth 2019;7(10):e13935) doi:10.2196/13935

KEYWORDS

preconception; subfertility; IVF treatment; pregnancy; cost-effectiveness

Introduction

Since the pioneer work of Edwards and Steptoe, in vitro fertilization (IVF) has become an indelible technology in modern era. Although the ongoing pregnancy rate after IVF treatment has tremendously increased [1], subfertility remains a worldwide problem affecting approximately 12% of couples of reproductive age [2]. In addition to the medical causes of subfertility, poor nutrition and lifestyle behaviors can impair fertility as well [3]. The mobile health (mHealth) coaching program Smarter Pregnancy [4,5] was developed to motivate (sub)fertile couples to adopt healthy nutrition and lifestyle behaviors. In a survey among (sub)fertile couples and a primary analysis of a randomized controlled trial (RCT) among couples with an IVF treatment indication, we showed that online coaching of

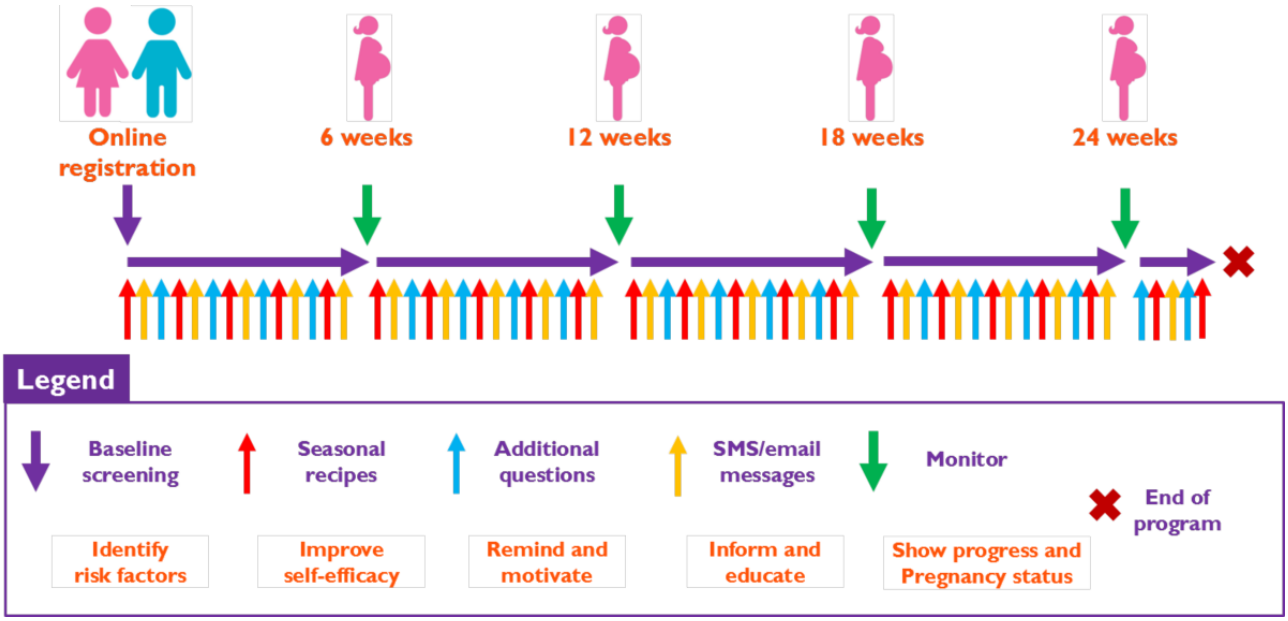
participants resulted in significant improvements of their nutrition and lifestyle behaviors [6,7]. Moreover, our survey also showed that improvements in nutritional behavior lead to an increase in ongoing pregnancy rates in fertile and subfertile couples with and without IVF treatment [8]. The health care and societal costs of IVF treatment are substantial [9], and we believe that many costs can be saved when a healthy lifestyle is adopted. Here, we aim to assess the cost-effectiveness of the use of this mHealth program compared to usual care in subfertile women preceding their first IVF treatment.

Methods

Study Population

The data were derived from a modelled study population consisting of subfertile women undergoing their first IVF treatment at the Erasmus MC, University Medical Center Rotterdam, the Netherlands. The data of the RCT were used to model nutrition and lifestyle behaviors. In this RCT, participants were randomly assigned to the intervention or control group. Participants of the intervention group received the complete coaching program and were coached on a maximum of five nutrition and lifestyle behaviors: vegetable, fruit, and folic acid supplement intake, smoking, and alcohol consumption. Participants of the control group only received a diminished version of the program. At several time points, all participants were asked to fill out questionnaires about their nutrition and lifestyle behavior. In this way, change in behavior could be measured. Participants of the RCT started the program at a maximum of 2 months before start of their IVF treatment, and the program lasted for a period of 24 weeks (Figure 1). The study protocol and primary results of the RCT on the improvement of these behaviors have been published elsewhere [6,10]. In brief, participants in the intervention group showed a significantly larger improvement in inadequate behavior compared to the control group [6].

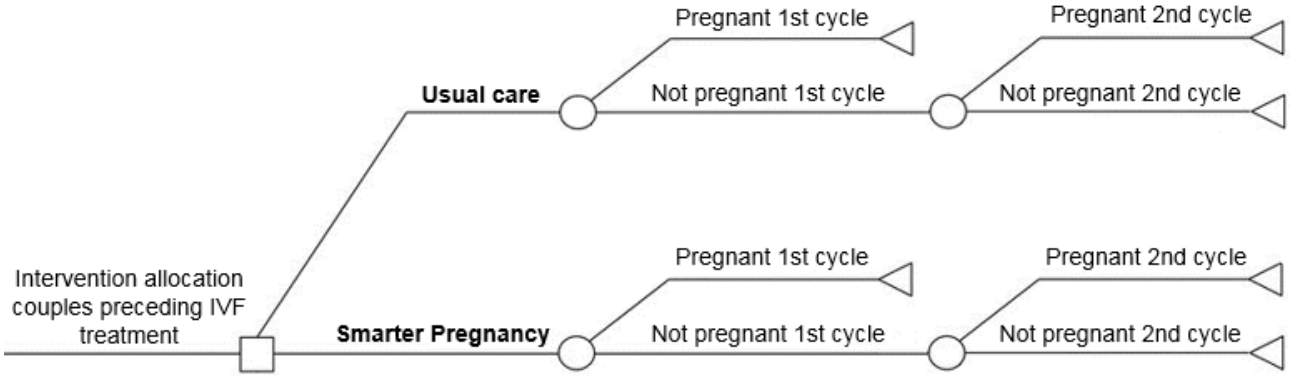
Figure 1. Overview of both the intervention and control groups during their enrollment in the Smarter Pregnancy randomized controlled trial. Adapted from van Dijk et al [<xref ref-type="bibr" rid="19ref10">10</xref>]. SMS: short message service.



Model

A decision tree model was constructed using Microsoft Excel (version 2010; Microsoft Corporation, Redmond, Washington) to assess the incremental ongoing pregnancies following the first IVF cycle and the costs of the mHealth program as compared to usual care (Figure 2). Ongoing pregnancy was defined as a vital pregnancy at 12 weeks of gestation. Women of subfertile couples who underwent their first IVF treatment in 2015 entered the model (n=793). A second IVF cycle was started if the first cycle did not result in an ongoing pregnancy. Pregnancy outcome following the second IVF cycle was the endpoint of the model. This short-term evaluation should therefore be considered a first indication of cost-effectiveness of the mHealth program.

Figure 2. Decision tree model. IVF: in vitro fertilization.



Model Scenarios

The usual care scenario reflects usual IVF treatment in the Netherlands. We assumed that all women received an elective single embryo transfer and that pregnancy rates in usual care are 33% for the first IVF cycle and 23% for the second cycle [11]. We furthermore assumed that all women in the intervention scenario were offered the mHealth program (100% coverage). This program was not offered in the usual care scenario (0% coverage). The intervention adherence rate was set at 70%, based on RCT data in which 70% of participants in the intervention group completed the coaching [6].

Model Parameters

Analyses were performed from a health care and societal perspective. The health care perspective includes costs related to the mHealth program [12], all costs associated with IVF treatment (eg, laboratory and hospital costs), and other relevant health care costs (eg, general practitioner visits). The societal perspective includes all health care costs plus costs outside the health care sector (eg, costs due to absence at work). The model parameters, including their distributions and sources, are reported in Table 1. Ongoing pregnancy rates after the first and second IVF cycle for the Smarter Pregnancy scenario and the usual care scenario were based on our previous study in the same setting [8] and others [13]. A detailed description of the cost calculations has been provided by Fiddelers et al [9]. All costs were expressed in euros (€) for the reference year 2016 based on the Dutch price index [14].

Table 1. Model input parameters.

Input parameter	Deterministic value	Probabilistic distribution	Source
IVF[a] costs (per cycle), €			
Hospital costs			
Hormone stimulation - medication	1580	Fixed	Fiddelers et al [9]
Hormone stimulation - hospital care	331	Fixed	Fiddelers et al [9]
Ovum pick-up	596	Fixed	Fiddelers et al [9]
Lab	1339	Fixed	Fiddelers et al [9]
Embryo transfer	316	Fixed	Fiddelers et al [9]
Other	295	Gamma	Fiddelers et al [9]
Other health care costs			
General practitioner	3	Gamma	Fiddelers et al [9]
Other	13	Gamma	Fiddelers et al [9]
Costs outside health care[b]			
Sick leave	569	Gamma	Fiddelers et al [9]
Leave of absence	141		Fiddelers et al [9]
Loss of leisure time	73	Gamma	Fiddelers et al [9]
Out of pocket expenditures	77	Gamma	Fiddelers et al [9]
Informal care	32	Gamma	Fiddelers et al [9]
Other	22	Gamma	Fiddelers et al [9]
Intervention costs, €			
Smarter Pregnancy program costs	61[c]	Gamma	Luyendijk [12]
Lifestyle costs[b], €			
Folic acid supplement use	64	Fixed	Luyendijk [12]
Healthy nutrition	113	Fixed	Luyendijk [12]
Smoking	1,223	Fixed	Based on data from [15] and [16]
Alcohol consumption	913	Fixed	Based on data from [17] and [18]
Pregnancy rates usual care			
First IVF cycle	0.329	Beta	Based on Wade et al [11]
Second IVF cycle	0.229	Beta	Based on Wade et al [11]
Pregnancy rate intervention			
First IVF cycle - 65% increase	0.543	Beta	Based on Twigt et al [8]
Second IVF cycle - 65% increase	0.443	Beta	Based on Twigt et al [8]

[a]IVF: in vitro fertilization.

[b]Only included in the analysis from a societal perspective. We assumed that the participants who smoke use 10 cigarettes per day (average daily use of smokers in the Netherlands) and that alcohol consumers drink one alcoholic beverage per day.

[c]Based on the annual tariff. This is considered to be an indication for the actual costs, which mainly consist of maintenance, insurance, overhead, and text messages.

Cost-Effectiveness Analysis

The primary effect outcome measure was expressed as the number of ongoing pregnancies after two IVF cycles. Incremental cost-effectiveness ratios (ICERs) from health care and societal perspectives were calculated by dividing the difference in costs between the Smarter Pregnancy scenario and the usual care scenario by the difference in the number of ongoing pregnancies in both scenarios. The ICER represents the estimated costs of one additional ongoing pregnancy.

A probabilistic sensitivity analysis was performed to consider the uncertainty surrounding the point estimates of the model input parameters. Probabilistic distributions were assigned to the parameters (Table 1). Thereafter, 1000 model iterations were performed by drawing random values from the distributions

assigned to the input parameters. We calculated the average costs and ongoing pregnancies by averaging these 1000 iterations. We performed deterministic sensitivity analyses to investigate the impact of changing several key parameters of the model: the coverage and adherence rate of the mHealth program and the chance of an ongoing pregnancy following the use of this program.

Results

Based on our model including 793 women, the mHealth scenario resulted in 369 pregnancies (47%; 95% CI 317-422) and the usual care scenario resulted in 283 pregnancies (36%; 95% CI 209-363) after two IVF cycles (Figure 1). The average health care costs for the mHealth and the usual care scenario were €6,008,500 (95% CI 5,671,000-6,505,000) and €6,214,800 (95% CI 5,839,500-6,730,300), respectively. The average societal costs for the mHealth and the usual care scenarios were €7,492,400 (95%CI 6,821,300;8,369,400) and €7,762,400 (95%CI 7,008,500-8,716,800), respectively (Figure 3). The ICERs from health care and societal perspectives per additional ongoing pregnancy equaled –€2250 (95%CI –3030 to –760) and –€3050 (95% CI –3960 to –540), respectively. Figure 4 shows that almost all ICERs are located in the southeast quadrant of the cost-effectiveness plane, indicating that the use of the mHealth program is cost saving.

The sensitivity analyses (Table 2) showed that the mHealth program is cost saving, on an average, but the uncertainty surrounding the ICERs increases when the intervention is less effective due to a lower compliance and ongoing pregnancy rate. For example, use of the mHealth program should increase the ongoing pregnancy rate by at least 51% for it to be cost saving compared to usual care when a 70% adherence rate is assumed. Otherwise, given an increased pregnancy rate of 65%, the compliance to Smarter Pregnancy should be at least 49% for it to remain cost saving.

Figure 3. Costs and effects (ongoing pregnancy rate) of the mobile health coaching program Smarter Pregnancy (intervention) and usual care, categorized as per health care and societal perspectives.

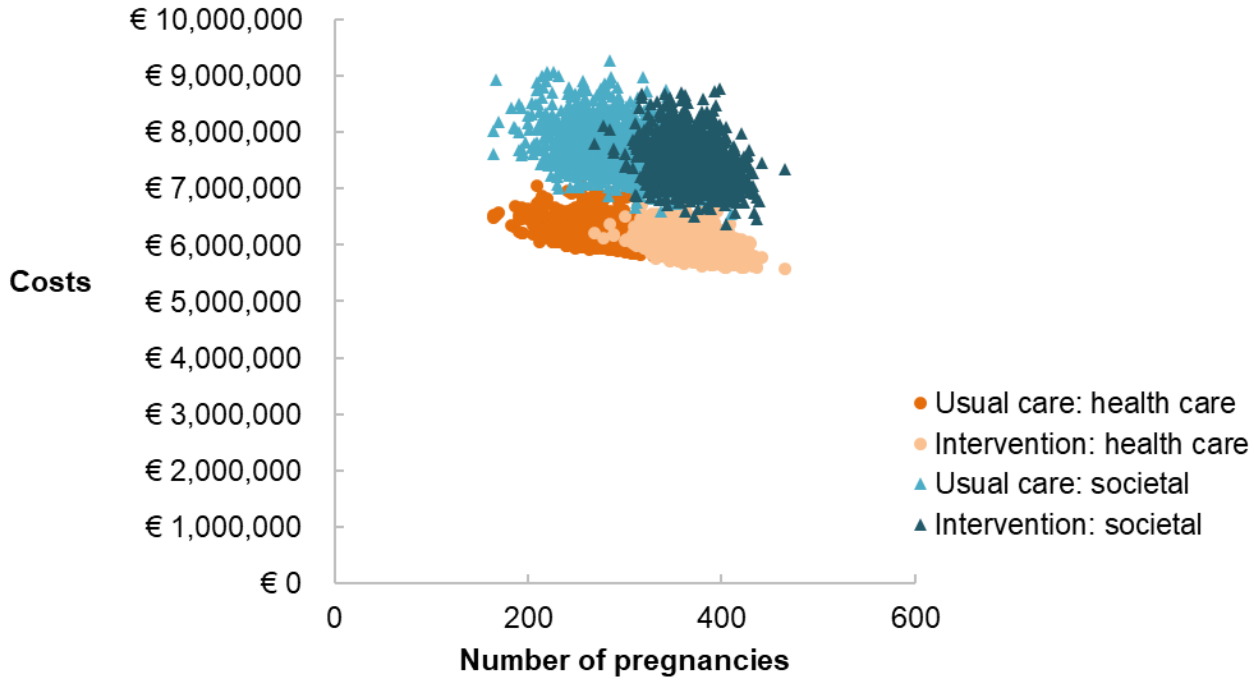


Figure 4. Incremental cost-effectiveness ratios generated by 1000 model simulations, categorized as per health care and societal perspectives.

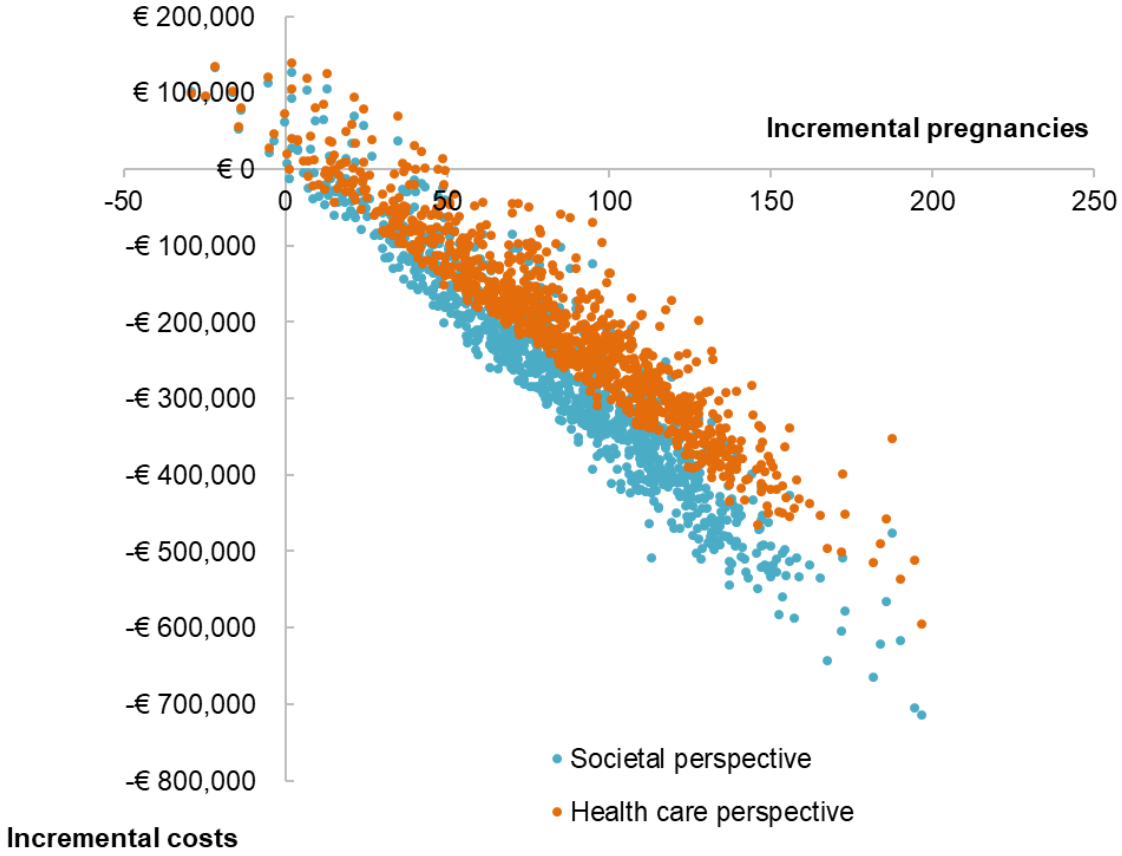


Table 2. Results of the sensitivity analyses.

Result	Mean number of incremental pregnancies	Mean incremental societal costs, €	Mean ICER[a] societal perspective (95%CI), €
Main analysis[b]	86	–270,000	–3050 (–3960 to –540)
Sensitivity analyses			
85% intervention compliance	105	–340,200	–3210 (–3960 to –1630)
55% intervention compliance	63	–192,000	–2840 (–3920 to –120)
45% increase in pregnancy rate (0.477)	64	–186,300	–3070 (–5610 to 1620)
25% increase in pregnancy rate (0.411)	40	–98,300	–2300 (–9610 to 9520)
70% intervention coverage	62	–187,600	–2840 (–3930 to –540)
85% intervention coverage	74	–227,400	–2850 (–3900 to –710)
Worst-case scenario[c]	21	–37,300	–1270 (–20,900 to 13,200)
Best-case scenario[d]	123	–408,900	–3600 (–3900 to –1850)

[a]ICER: incremental cost-effectiveness ratio.

[b]100% intervention coverage, 70% intervention compliance, 65% increase in pregnancy rate.

[c]70% intervention coverage, 55% intervention compliance, 25% increase in pregnancy rate.

[d]100% intervention coverage, 100% intervention compliance, 65% increase in pregnancy rate.

Discussion

Principal Findings

This model-based study, determining the estimates of available data, showed that the use of the mHealth program would result in 86 additional pregnancies and a reduction of €270,000 compared to usual care after two IVF cycles, resulting in an ICER of –€3050 per additional ongoing pregnancy. Sensitivity analyses showed that the use of this mHealth program is cost saving when the ongoing pregnancy rate increases to at least 51% after two cycles of single embryo-transfer IVF treatment.

Strengths and Limitations

A strength of our model is the combined use of evidence-based data of the population, clinical effectiveness, compliance, and costs to support decision making. Although model parameters would ideally be based on meta-analyses or larger datasets, these were unavailable. Since the Smarter Pregnancy RCT is ongoing, assumptions regarding ongoing pregnancy rates had to be made based on our previous data. In economic evaluations, a time horizon that is long enough to capture all relevant costs and effects is preferred [19]. Our study was limited to two IVF cycles, which may be relatively short. However, as the endpoint of our study was to assess the incremental ongoing pregnancy rate, other costs and long-term reproductive and health outcomes were not considered.

We evaluated single embryo transfers only, because in the Netherlands, this is the most common IVF strategy. Therefore, costs and ongoing pregnancy rates of other IVF strategies will be different [13].

Comparison With Prior Work

Several studies have investigated the effectiveness of nutrition and lifestyle interventions preceding fertility treatment. However, most of these studies focus on specific patient groups such as obese or anovulatory women [20,21]. In accordance with our findings, the study by Van Oers et al [22] showed that lifestyle intervention preceding fertility treatment was cost-effective in terms of achieving an ongoing pregnancy within 24 months.

The difference in average societal costs and health care costs was relatively small, indicating that the addition of the non-healthcare costs had no substantial impact on the ICER. Because nutrition and lifestyle interventions in preconception care have relatively low additional budget impact, we expect that the chance that the mHealth program is not cost-effective would be low [23].

Conclusions

Our results show that the mHealth coaching program Smarter Pregnancy is potentially cost saving for subfertile couples preceding their first IVF treatment. Although our results are promising, our model requires further validation based on actual data on ongoing pregnancy rates from the Smarter Pregnancy RCT in order to establish the relative cost-effectiveness of the mHealth program with greater certainty. Implementation of this mHealth program in routine preconception care of subfertile couples should be seriously considered, given the relatively low intervention costs and promising cost-effectiveness estimates.

Acknowledgments

This study was funded by a grant from ZonMW; the Netherlands Organization for Health Research and Development (project number 209040003), and the Department of Obstetrics and Gynecology of the Erasmus MC, University Medical Center, Rotterdam, the Netherlands. The sponsor had no role in the analysis or preparation of the manuscript.

Conflicts of Interest

RS-T is CEO of eHealth Care Solutions (since 2013). JL has received grants from Ferring BV, Dutch Heart Association, ZonMw, Danone, Euroscreen, Anshlabs, and Titus Healthcare outside the submitted work. No other disclosures are reported.

References

1. Kushnir VA, Barad DH, Albertini DF, Darmon SK, Gleicher N. Systematic review of worldwide trends in assisted reproductive technology 2004-2013. Reprod Biol Endocrinol 2017 Jan 10;15(1):6 [FREE Full text] [doi: 10.1186/s12958-016-0225-2] [Medline: 28069012]
2. Inhorn MC, Patrizio P. Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century. Hum Reprod Update 2015;21(4):411-426 [FREE Full text] [doi: 10.1093/humupd/dmv016] [Medline: 25801630]
3. Homan GF, Davies M, Norman R. The impact of lifestyle factors on reproductive performance in the general population and those undergoing infertility treatment: a review. Hum Reprod Update 2007;13(3):209-223. [doi: 10.1093/humupd/dml056] [Medline: 17208948]
4. Smarter Pregnancy: Personal coaching for healthy food- and lifestyle habits. URL: https://www.smarterpregnancy.co.uk/research/ [accessed 2018-12-29] [WebCite Cache ID 751h3BVP3]
5. Slimmer Zwanger: Persoonlijke coaching voor gezonde voedings- en leefstijlgewoonten. URL: https://www.slimmerzwangeronderzoek.nl/nl/ [accessed 2018-12-29] [WebCite Cache ID 751hektYe]

6. Oostingh E, Koster M, van Dijk M, Willemsen S, Steegers E, Laven J. Improvement of Periconception nutrition and lifestyle behaviors using the mHealth program: a randomized controlled trial. 2018 Mar 25 Presented at: Society for Reproductive Investigation; March 25, 2018; Orlando p. 2018.
7. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth Platform for Pregnancy on Nutrition and Lifestyle of the Reproductive Population: A Survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
8. Twigt JM, Bolhuis MEC, Steegers EAP, Hammiche F, van IWG, Laven JSE, et al. The preconception diet is associated with the chance of ongoing pregnancy in women undergoing IVF/ICSI treatment. Hum Reprod 2012 Aug;27(8):2526-2531 [FREE Full text] [doi: 10.1093/humrep/des157] [Medline: 22593431]
9. Fiddelers AAA, van Montfoort APA, Dirksen CD, Dumoulin JCM, Land JA, Dunselman GAJ, et al. Single versus double embryo transfer: cost-effectiveness analysis alongside a randomized clinical trial. Hum Reprod 2006 Aug;21(8):2090-2097. [doi: 10.1093/humrep/del112] [Medline: 16613886]
10. van Dijk MR, Oostingh EC, Koster MPH, Willemsen SP, Laven JSE, Steegers-Theunissen RPM. The use of the mHealth program Smarter Pregnancy in preconception care: rationale, study design and data collection of a randomized controlled trial. BMC Pregnancy Childbirth 2017 Jan 26;17(1):46 [FREE Full text] [doi: 10.1186/s12884-017-1228-5] [Medline: 28125970]
11. Wade JJ, MacLachlan V, Kovacs G. The success rate of IVF has significantly improved over the last decade. Aust N Z J Obstet Gynaecol 2015 Oct;55(5):473-476. [doi: 10.1111/ajo.12356] [Medline: 26174052]
12. Luyendijk M. Cost-effectiveness of an e-Health intervention to improve unhealthy lifestyle behaviors in subfertile women. 2013 Aug 22. URL: https://thesis.eur.nl/pub/15868/ [accessed 2019-09-02]
13. Fiddelers AAA, Dirksen CD, Dumoulin JCM, van Montfoort APA, Land JA, Janssen JM, et al. Cost-effectiveness of seven IVF strategies: results of a Markov decision-analytic model. Hum Reprod 2009 Jul;24(7):1648-1655. [doi: 10.1093/humrep/dep041] [Medline: 19318703]
14. Centraal Bureau voor de Statistiek. StatLine. 2017. Consumentenprijzen; prijsindex 2015=100 Gewijzigd op: 6 augustus 2019 URL: https://opendata.cbs.nl/#/CBS/nl/dataset/83131NED/table?dl=5400 [accessed 2017-01-01] [WebCite Cache ID 751hhvT33]
15. Stichting stop bewust. Wat kost roken URL: https://ikstop.nl/wat-kost-roken [accessed 2018-12-29] [WebCite Cache ID 751hjc14L]
16. Stichting stop bewust. Roken en statistiek URL: https://ikstop.nl/roken-en-statistiek [accessed 2018-12-29] [WebCite Cache ID 751hlD86r]
17. Trimbos instituut. Cijfers alcohol URL: https://www.trimbos.nl/kerncijfers/cijfers-alcohol [accessed 2018-12-29] [WebCite Cache ID 751iCTXWQ]
18. Jellinek. Informatie over alcohol en drugs URL: https://www.jellinek.nl/informatie-over-alcohol-drugs/alcohol/alcohol-basisinfo/soorten-en-prijzen-alcohol [accessed 2018-12-29] [WebCite Cache ID 751hoN05p]
19. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford, UK: Oxford University Press; 2006:978-019.
20. Sim KA, Partridge SR, Sainsbury A. Does weight loss in overweight or obese women improve fertility treatment outcomes? A systematic review. Obes Rev 2014 Oct;15(10):839-850. [doi: 10.1111/obr.12217] [Medline: 25132280]
21. Legro RS, Dodson WC, Kunselman AR, Stetter CM, Kris-Etherton PM, Williams NI, et al. Benefit of Delayed Fertility Therapy With Preconception Weight Loss Over Immediate Therapy in Obese Women With PCOS. J Clin Endocrinol Metab 2016 Dec;101(7):2658-2666 [FREE Full text] [doi: 10.1210/jc.2016-1659] [Medline: 27172435]
22. van Oers AM, Mutsaerts MAQ, Burggraaff JM, Kuchenbecker WKH, Perquin DAM, Koks CAM, LIFEstyle study group. Cost-effectiveness analysis of lifestyle intervention in obese infertile women. Hum Reprod 2017 Dec 01;32(7):1418-1426. [doi: 10.1093/humrep/dex092] [Medline: 28486704]
23. Naberhuis J, Hunt VN, Bell JD, Partridge JR, Goates S, Nuijten MJC. Health care costs matter: a review of nutrition economics – is there a role for nutritional support to reduce the cost of medical health care? NDS 2017 Aug 01;Volume 9:55-62 [FREE Full text] [doi: 10.2147/nds.s126232]

Abbreviations

ICER: incremental cost-effectiveness ratio
IVF: in vitro fertilization
mHealth: mobile health
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 06.03.19; peer-reviewed by R Norman, A van Oers, D Rodger; comments to author 16.05.19; revised version received 29.06.19; accepted 28.07.19; published 23.10.19.

Please cite as:
Oostingh EC, Ophuis RH, Koster MPH, Polinder S, Lingsma HF, Laven JSE, Steegers-Theunissen RPM
Mobile Health Coaching on Nutrition and Lifestyle Behaviors for Subfertile Couples Using the Smarter Pregnancy Program: Model-Based Cost-Effectiveness Analysis
JMIR Mhealth Uhealth 2019;7(10):e13935
URL: http://mhealth.jmir.org/2019/10/e13935/
doi:10.2196/13935
PMID:31647476

Original Paper

Factors Influencing Exercise Engagement When Using Activity Trackers: Nonrandomized Pilot Study

Amanda Jayne Centi[1], PhD; Mursal Atif[1], BA; Sara Bersche Golas[1], MA; Ramin Mohammadi[2], BS; Sagar Kamarthi[2], PhD; Stephen Agboola[1,3,4], MD, MPH; Joseph C Kvedar[1,3,4], MD; Kamal Jethwani[1,3,4], MD, MPH

[1]Pivot Labs, Partners Healthcare, Boston, MA, United States

[2]Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, United States

[3]Department of Dermatology, Massachusetts General Hospital, Boston, MA, United States

[4]Harvard Medical School, Harvard University, Boston, MA, United States

Corresponding Author:
Amanda Jayne Centi, PhD
Pivot Labs
Partners Healthcare
25 New Chardon St, Ste 300
3rd Floor
Boston, MA, 02114
United States
Phone: 1 6177242158
Fax: 1 6177267530
Email: acenti@partners.org

Abstract

Background: It is well reported that tracking physical activity can lead to sustained exercise routines, which can decrease disease risk. However, most stop using trackers within a couple months of initial use. The reasons people stop using activity trackers can be varied and personal. Understanding the reasons for discontinued use could lead to greater acceptance of tracking and more regular exercise engagement.

Objective: The aim of this study was to determine the individualistic reasons for nonengagement with activity trackers.

Methods: Overweight and obese participants (n=30) were enrolled and allowed to choose an activity tracker of their choice to use for 9 weeks. Questionnaires were administered at the beginning and end of the study to collect data on their technology use, as well as social, physiological, and psychological attributes that may influence tracker use. Closeout interviews were also conducted to further identify individual influencers and attributes. In addition, daily steps were collected from the activity tracker.

Results: The results of the study indicate that participants typically valued the knowledge of their activity level the activity tracker provided, but it was not a sufficient motivator to overcome personal barriers to maintain or increase exercise engagement. Participants identified as extrinsically motivated were more influenced by wearing an activity tracker than those who were intrinsically motivated. During the study, participants who reported either owning multiple technology devices or knowing someone who used multiple devices were more likely to remain engaged with their activity tracker.

Conclusions: This study lays the foundation for developing a smart app that could promote individual engagement with activity trackers.

(JMIR Mhealth Uhealth 2019;7(10):e11603) doi:10.2196/11603

KEYWORDS

activity trackers; exercise; engagement

Introduction

Background

Despite the well-recognized benefits of physical activity, millions of people are physically inactive, and the prevalence of physical inactivity is increasing, placing people at a greater risk for obesity and many other cardio metabolic disorders [1,2]. In 2016, physical inactivity was reported as the fourth leading cause of mortality [3]. Yet, as we age, we spend more time physically inactive. It is estimated that Americans aged 20 to 29 years spend 55% of their time inactive, whereas those aged 70 to 79 years spend 67% of their time inactive [4]. Currently, there is much research into methods to not only increase physical activity but also forge sustainable physical activity patterns. Technology plays an important and promising role in personal activity tracking. Wearable activity trackers have been championed as powerful personalized health management tools because of their "low cost, wide reach, and apparent effectiveness," and the commercial market for such devices is large and expanding [5,6]. In 2013, 1 out of every 5 US adults surveyed reported using "some form of technology to track their health data," including medical devices, mobile phone apps, or Web-based tools [7]. Individual consumers and health care providers recognize the potential benefits of wearable activity trackers, which may be used to monitor many health indicators, including diet, physical activity, and sleep [8]. In addition, activity trackers can be beneficial in aiding with chronic disease management by promoting behavioral health changes encouraged by health care providers, such as increasing physical activity [5]. It has been reported that the use of technology to monitor physical activity was associated with higher levels of activity [8]. However, the potential benefits derived from the use of activity trackers are challenged by the limited and transient adoption of these devices, which requires sustained use to achieve their intended effect [8]. Although it has been well documented that most users lose interest in using trackers not long after purchase, specific reasons for this remain to be deciphered [9]. Currently, little is known about the factors associated with the adoption and sustained use of activity trackers or the barriers that limit the effectiveness of these devices in people's efforts to increase physical activity and improve health. Perceived barriers to physical activity have been previously studied, and it has been demonstrated that motivational factors are associated with physical activity level [1]. Lack of time, fatigue, and a dislike for exercise are some of the barriers that reflect a lack of motivation to engage in physical activity [1]. If the characteristics and patterns that lead to a person becoming disengaged with his or her tracker are also better understood, interventions may be created to maintain engagement. However, little is known about when a person may disengage with his or her tracker and what personal characteristics may be influencing that decision.

Objective

Therefore, purpose of this study was to understand reasons for engagement and disengagement associated with the use of activity trackers in an overweight population. We chose an overweight population as our first use case, as the benefits of maintaining or even increasing physical activity in this population have been well established, whereas sustaining interventions are lacking [10-12].

Methods

Study Design

This was a 9-week, nonrandomized pilot study designed to explore activity tracker engagement patterns to guide and inform future work into the development of a predictive algorithm to facilitate user engagement with trackers.

Study Population and Setting

Participants (n=30) aged 18 years and older, with a body mass index of 25 kg/m^2 or greater, were recruited from a Massachusetts General Hospital clinic. After screening into and consenting to the study, participants were directed to the study website (wellocracystudy.org), where they were asked to read through information regarding the study and types of activity trackers available to use for the 9-week study and keep after study completion. Participants could choose from the following FitBit activity trackers: Charge, Flex, One, or Zip. After choosing a device, study staff assisted with setup as necessary.

Data Collection

Participants were asked to wear the activity tracker continuously during the 9-week study period. The first week was used as a run-in period to determine the participant's baseline average daily steps. A step goal was then set at 10% above this average. Each participant's step goal was unique to the participant on the basis of the participant's activity during week 1. Participants were sent a short message service text message to change their step goal to this amount for the remaining 8 weeks. During the 8 weeks, minimal contact was made by study staff with participants so that their habits of using the activity tracker could be assessed. Study staff would download the data weekly from the FitBit application programming interface. If it was found that no data were being collected from the activity tracker, study staff reached out to the participant and attempted to provide support on pairing the device and uploading data in the FitBit app. If a participant indicated not using the device, reasons for nonuse and engagement were documented and explored in depth with a trained staff member. After the study, participants completed a closeout survey either on the Web or in paper format and underwent a phone interview with a neuropsychologist to gather information about their experiences during the study, which was transcribed for analysis. Baseline and closeout questionnaires included questions about their thoughts about and perceived barriers to exercise and activity—Behavioral Regulation in Exercise Questionnaire (BREQ-2) [13] and Barriers to Being Active (BBA) [14]), Prochaska's Stage of Change [15], and general health questions (Patient-Reported Outcomes Measurement Information System, PROMIS Global-10) [16]. As episodes of severe depression can impact physical activity in ways beyond the scope of this intervention, all participants underwent a screening—Patient Health Questionnaire (PHQ-8) [17]. In addition, technology use, ownership, and demographic data were collected at baseline only. The BREQ-2 is designed to gauge the extent to which people's reasons for exercise are internalized and

self-determined on the basis of the following categories: amotivation, external, introjected, identified, intrinsic. The BBA assesses whether participants gauge certain categories as reasons for inactivity, including energy, willpower, time, and resources. For each category, a score of 5 or greater would indicate that category as a substantial barrier to a person's ability to exercise. The PROMIS Global-10 is a 10-item survey that seeks to assess health care–related quality of life along 2 metrics, physical and mental health for all participants, and it was administered at enrollment and closeout. The PHQ-8 is a brief survey of a person's depression status, whereas Prochaska's Stage of Change assesses a person's state for changing current habits and behaviors. In addition to the survey questions, participants were asked to complete a poststudy phone interview with trained staff. Although the questions were open ended in general, the main purpose of these interviews was to decipher more specific reasons for how participants engaged with their activity tracker during the study period. These interviews were recorded and transcribed. Each interview lasted between 30 and 45 min in length.

Data Analysis

For analysis, participants were divided into 3 *a posteriori* "engagement groups" on the basis of the percent of days they met their step goal. Step data were collected and analyzed to determine average number of steps, amount of time participants wore their device, and the percent of days a person met his or her step goal. Each participant's step goal was determined by taking the participant's average number of steps from Week 1 and adding 10% to this number. A weekly average of steps per day was then calculated for each participant. Questionnaires were scored according to their standard practice. In instances where 1 of the questionnaires was skipped, values from the enrollment questionnaire were carried over for analysis purposes. These instances are indicated in the descriptions below. Qualitative data from closeout interviews were analyzed by trained staff in qualitative analysis who conducted a thematic analysis of the transcribed interviews for key patterns to the motives and barriers to using activity trackers. Means and SDs were used for continuous variables. Categorical variable percentages were calculated as percent of group total.

Results

At the beginning of the study, a total of 30 participants were enrolled. Among them, 21 participants completed both enrollment and closeout procedures as part of the study. The remaining 9 participants were lost to follow-up. Baseline characteristics of all enrolled participants (n=30) are summarized in Table 1.

Overall, the participant distribution comprised 60% female, 70% white, 47% employed, and 60% individuals with at least some education post-high school. As the distribution of engagement groups was skewed—Shapiro–Wilk P value≤.01; median (Q1, Q3)=37.3% (26.2%, 51.3%); range=0%-83.7%. Groups were defined by quartile: the bottom 2 quartiles= "Low engagement," the third quartile="Medium engagement," and the upper quartile="High engagement." Patients lost to follow-up were classified as "Nonengaged," as they were not engaged enough with the tracker to complete the 9-week study. Nonengaged participants were included in the study analysis to try to determine initial characteristics that may be identified before the study to keep similar future participants engaged during a follow-up study. Though Fisher Exact test revealed a statistical significance among engagement groups for Marital Status (P=.01), engagement groups were statistically similar on all other demographic variables. All but 3 participants chose to use the FitBit Charge model. A person in the high and medium groups used the FitBit One, whereas a person in the nonengaged group used the FitBit Zip.

Over the 8 weeks of data collection, the number of participants who met their weekly step goals was low. Overall, less than 50% of participants met their step goal each week (Table 2).

The number of days participants met their step goal over the study period ranged from 3 days to 38 days, with an average of 20.2. As there seemed to be a range in participants meeting their step goal, and we wanted to determine facilitators or barriers to meeting this goal, the Engagement Level categories described above were used to look for patterns and characteristics that could be used to profile how and why participants engaged in certain manners. Of the 30 participants enrolled, 28 completed the enrollment questionnaire, and 21 complete the closeout questionnaire. The patient Enrollment and Closeout Questionnaire results are summarized below.

Table 1. Participant demographics (N=30).

Variable	Value
Age (years), mean (SD)	48.96 (9.54)
Gender (male), n (%)	9 (30)
Body mass index at enrollment, mean (SD)	32.48 (4.59)
Range	25-41.2
Race, n (%)	
White	21 (70)
Nonwhite	9 (30)
Marital status, n (%)	
Married	8 (26.7)
Divorced/separated	8 (26.7)
Single (never married)	8 (26.7)
Living with partner	3 (10)
Widowed	1 (3.3)
No response	2 (6.7)
Education, n (%)	
12 years or completed high school or General Education Diploma	5 (16.7)
Some college	5 (16.7)
College graduate	9 (30)
Posthigh school	2 (6.7)
Postgraduate	2 (6.7)
Less than high school	3 (10)
Unknown	4 (13.3)
Employment status, n (%)	
Employed/self employed	15 (50)
Disabled	5 (16.7)
Unemployed	5 (16.7)
Student	1 (3.3)
Retired	1 (3.3)
Unknown	3 (10)

Table 2. Percent of participants meeting their step goal (based on week 1 data), by week, over the course of the study.

Week	Patients who met goal (%)
2	23
3	50
4	45
5	23
6	41
7	32
8	23
9	27

Friends and Family Who Track

During enrollment, participants were asked if they knew anyone who used trackers. We found that participants with friends or family who used trackers were more engaged over the course of the study. Though not significant (P=.09), less engaged participants were less likely than those who continued engagement to have friends or family members with trackers: 50% to 80% of those who continued in the study have friends, family, or both who track, whereas the same can be said for only 14% of the nonengaged.

Technology Ownership

Participants were asked whether they owned particular items of technology, such as a desktop computer, a laptop computer, or a tablet. Nonengaged participants were more likely to own only 1 device, whereas other engagement levels were more likely to own 2 or more devices (Figure 1, P=.18).

Stage of Change: Enrollment and Closeout

At enrollment, there were no differences in Stage of Change observed between groups (Figure 2, P=.67). At closeout (see Figure 3), there was a significant difference in Stage of Change between the groups (P=.04).

Figure 1. Device ownership, by engagement level.

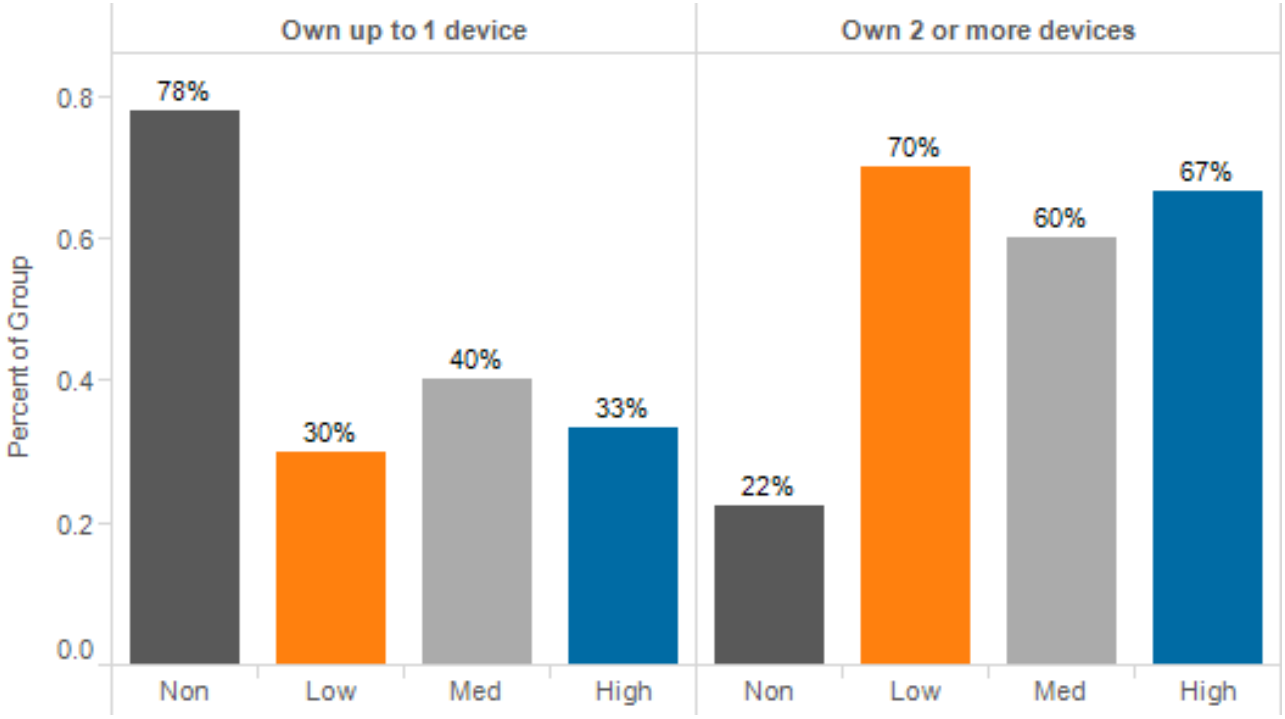


Figure 2. Stage of Change at enrollment.

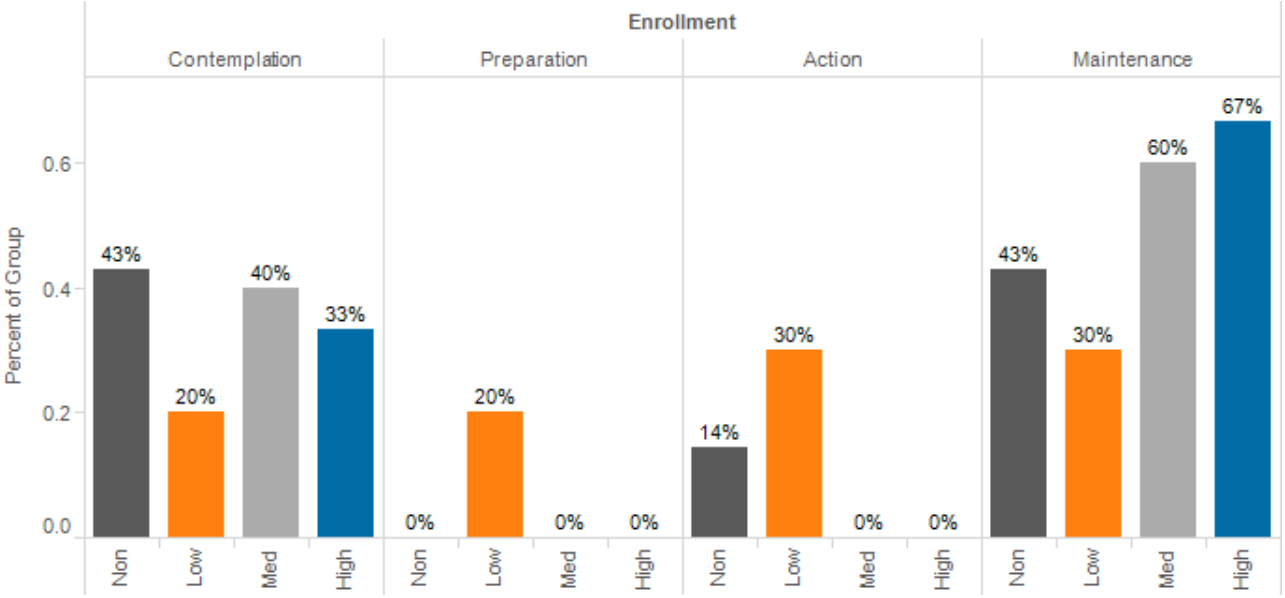
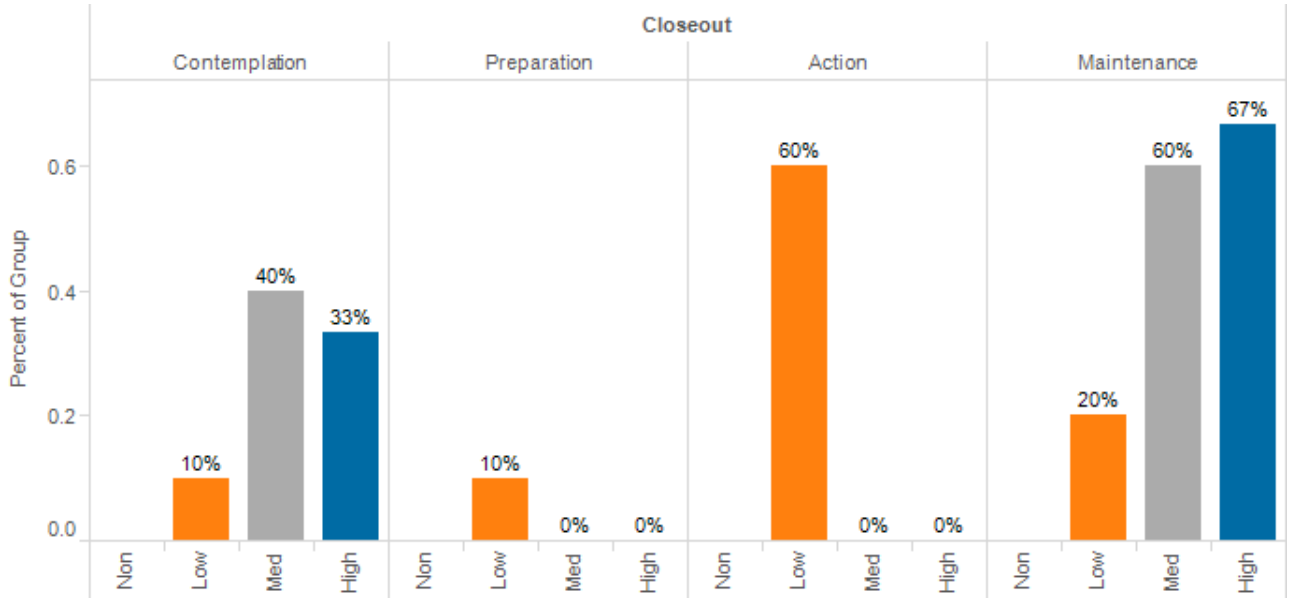


Figure 3. Stage of Change at closeout, by engagement level.



Patient Health Questionnaire-8: Enrollment and Closeout

There was no difference in PHQ-8 scores at enrollment (non=8±5.7, low=4.5±3.6, medium=5.2±5, and high=8.2±8) or closeout (low=6±5.3, medium=6.4±2.9, and high 6±6.4). Though not significant, the High-Engagement group saw a decrease in PHQ score (P=.36), whereas the low- and medium-Engagement groups both saw a slight increase (P=.39 and P=.50, respectively). Total scores of 10 or greater are indicative of depression, whereas scores of 20 or greater are indicative of severe depression. At enrollment, there were 5 participants who met the score for depression (non=2, low=1, medium=1, and high=2), whereas at closeout, there were 4 participants who did not (low=2, medium=1, high=1).

Behavioral Regulation in Exercise Questionnaire

Table 3 presents the percentage of each group who had each degree of self-determination as its highest BREQ category score at Enrollment and Closeout. Although most patients had "identified" or "intrinsic" as their highest score, no group had "amotivation" or "external" as its highest score.

Overall, the averages for less self-determined categories (Amotivation, External) increased, whereas the averages for more self-determined categories (Identified, Intrinsic) decreased from enrollment to closeout.

Table 3. Highest Behavioral Regulation in Exercise Questionnaire-2 Category score by group at enrollment and closeout (percentage of group).

Group	Enrollment classification				Closeout classification			
	Tied, %	Introjection, %	Identified, %	Intrinsic, %	Tied, %	Introjection, %	Identified, %	Intrinsic, %
Non	—[a]	14	29	57	—	—	—	—
Low	20	10	40	30	10	—	70	20
Medium	—	—	60	40	—	20	40	40
High	17	—	33	50	17	17	50	17

[a]Not applicable.

Barriers to Being Active

At enrollment, lack of willpower was the highest average category score for all 4 engagement groups, as well as the most frequent barrier (Figure 4). At closeout, lack of willpower still had the highest average category for the low- and medium-engagement groups, whereas the high-engagement group showed a decrease in this score, with lack of resources having the highest average. A similar pattern is seen with regards to percentage of each group for whom lack of willpower is a barrier (Figure 5).

Between groups at enrollment (ENR), lack of time was the only significantly different category among engagement groups (Table 4; P=.03). However, this difference did not remain at closeout (CLS). Within groups, there were no significant differences for any group or category, and most remained similar for each category, with the exception of lack of energy for the low engagement group and lack of time and willpower for the high-engagement group. Between and within groups P values are presented in Table 5 below.

Figure 4. Percent of group for whom category is a barrier, by engagement level at enrollment. ENR: enrollment; Grp: group.

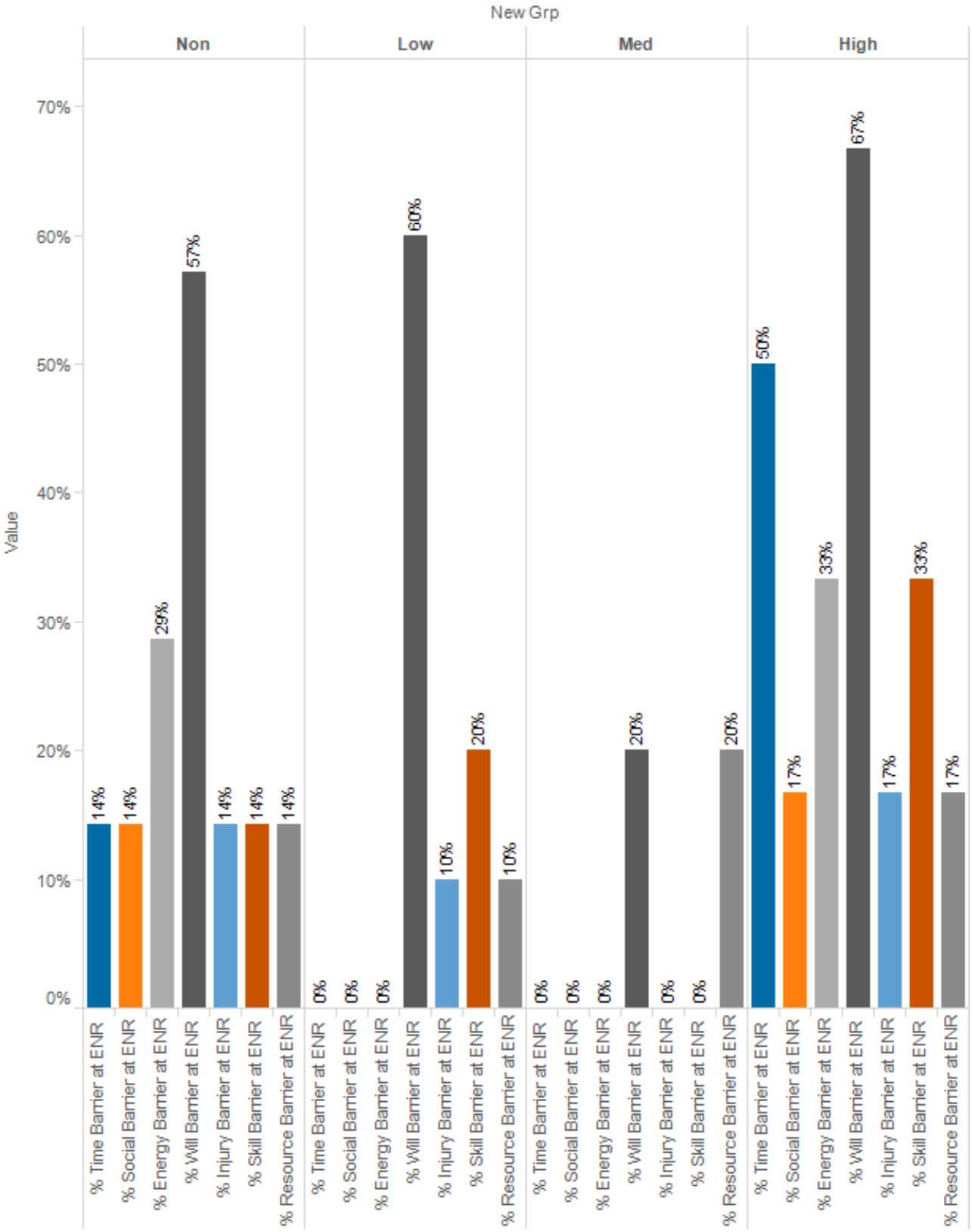


Figure 5. Percent of group for whom category is a barrier, by engagement level at closeout. CLS: closeout. Grp: group.

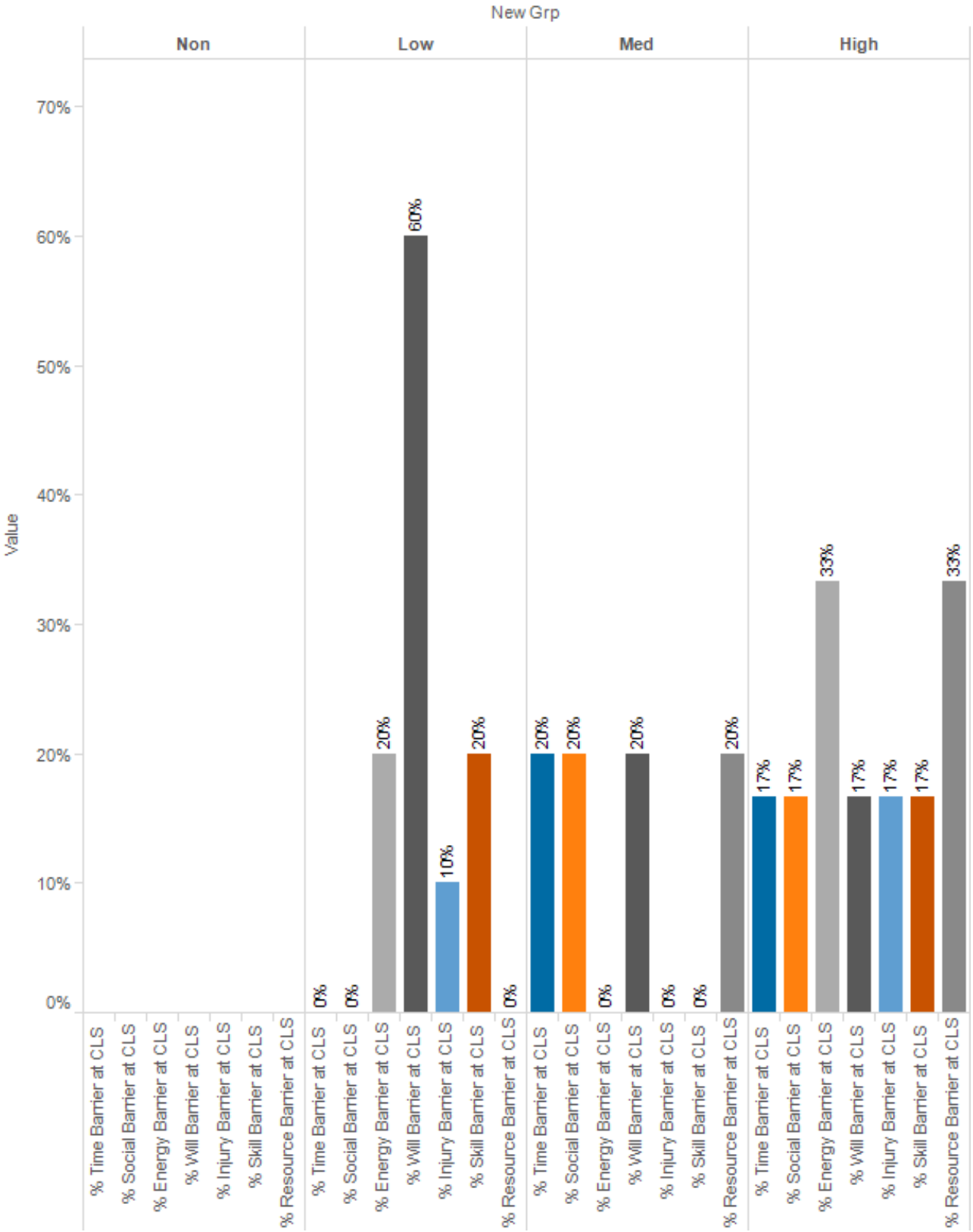


Table 4. *P* values for between (B) and within (W) group differences, by group, category, and time point (enrollment vs closeout).

Category	Enrollment (B)	Closeout (B)	Low (W)	Medium (W)	High (W)
Lack of time	*.03*[a]	.26	1.00	1.00	*.55*
Social barriers	.40	.26	1.00	1.00	1.00
Lack of energy	.10	.53	*.47*	1.00	1.00
Lack of willpower	.47	.18	1.00	1.00	*.24*
Fear of injury	1.00	1.00	1.00	1.00	1.00
Lack of skill	.68	.77	1.00	1.00	1.00
Lack of resources	1.00	.12	1.00	1.00	1.00

[a]Italics indicate significance.

Table 5. *P* values for between (B) and within (W) group differences, by group, category, and time point (enrollment vs closeout): category scores. Italics denote significance.

Category	Enrollment (B)	Closeout (B)	Low (W)	Medium (W)	High (W)
Lack of time	.43[a]	.96	.40[b]	.17	.18[c]
Social barriers	.19	.92	.66	.21	1.00
Lack of energy	.47[d]	.86	.29	.14	.62
Lack of willpower	.76	.59	.83	.70	.20
Fear of injury	.51	.40	.79	1.00	.71
Lack of skill	.36	.48	1.00	1.00	.59
Lack of resources	.13	*.03*	.52	.07	1.00

[a]Kruskal-Wallis test.

[b]Mann Whitney *U* test.

[c]Paired *t* test.

[d]One-way analysis of variance.

For category scores, between and within group comparisons (Table 5) show the groups to be similar across categories and timepoints, with the exception of between groups scores for lack of resources at closeout.

General Perceived Health

Scores from the PROMIS Global-10 are reported in Table 6. Scores are presented for Physical (Global Physical Health, GPH) and Mental (Global Mental Health) health scores. The participants in the high-engagement group were the only participants in whom an increase in GPH was observed from enrollment to closeout (*P*=.04)

Closeout interviews were conducted on 15 participants. From these interviews, common themes that helped and hindered tracker use were compiled. Common influencers to high or low tracker use included participant's health status, pain level, weather, emotional state, and daily agenda/routine. Other barriers to use included a preference for more sedentary activities, insufficient space to exercise, and difficulty starting a new routine. Motivators that helped participants increase their activity were categorized as either extrinsic or intrinsic. Extrinsic motivators included personal/internal goals (lose weight, improve health and energy level) and externally imposed goals (tracker goal and rewards, social comparison), whereas the largest intrinsic motivator reported was an enjoyment of being active. From the closeout interviews, 4 criteria were identified to more likely trigger a change in activity level and routine. These included the following: (1) a routine that allowed for exercise without significant barriers, (2) an extrinsic motivator to pursue an activity, (3) a clear and specific personal activity goal, and (4) an extrinsic accountability mechanism (personal trainer, workout buddy, and activity tracker).

Table 6. Global Physical Health/Global Mental Health Scores at enrollment (n=27) and closeout (n=20), by engagement level. Italics indicate significance.

Category	Engagement level				Between-group *P* value
	Non (n=7)	Low (n=10)	Medium (n=5)	High (n=5)	
Global Physical Health, mean (SD)					
Enrollment	36.2 (8.05)	41.8 (7.61)	36.9 (8.92)	37.5 (8.33)	.49
Closeout	—[a]	42.1 (6.88)	43.7 (7.83)	47.0 (6.69)	.45
Within-group *P* value	—	.89	.15	*.04*	—
Global Mental Health, mean (SD)					
Enrollment	43.5 (5.22)	45.0 (5.12)	51.3 (3.77)	42.0 (7.76)	.06
Closeout	—	43.0 (4.77)	48.3 (2.50)	45.9 (5.13)	.11
Within-group *P* value	—	.15	.18	.29	—

[a]Not applicable.

Discussion

To our knowledge, this is one of the first studies to take an in-depth look at the reasons participants choose to engage with activity trackers and work to develop attributes that can be identified to better predict a person's engagement. In general, we were able to determine that those who reported owning more devices were more likely to be engaged with their activity tracker than those with less technology ownership. In addition, those who reported knowing someone who used a tracker were also more likely to remain engaged with their activity tracker. Both of these findings would point to a need for participants to feel a level of comfort with either using the device or having support to help them with the device as a method of increasing engagement. The findings may also point to a need for very simple activity trackers to be able to reach a wider range of people. If there is a low technology barrier for a person to overcome, the person may not need as much experience or support to sustain use. Ease of use was a reason for choosing the FitBit platform for this study, and it has been noted in other studies as well [18]. These reasons can be added to previous reports, which indicate that former users of activity trackers indicated learning all they could, and former users specified trackers not helping them achieve goals as reasons for disengaging from their tracker [19]. A common theme we heard from participants at closeout was a benefit of receiving an actual measure of their physical activity. This is similar to what has been observed in previous reports. In a study on acceptance of commercially available wearable activity trackers among adults over 50, participants reported that the most beneficial aspect of the using an activity tracker was increase in self-awareness of activity levels [3]. In addition, a recently published study noted that over 81% of people using trackers thought it made them more physically active [19]. If we are able to manipulate this self-awareness, it may be another useful tool for keeping participants engaged, at least for a certain amount of time. However, from this study and previous studies, there also seems to be a need for continuation of learning to sustain use [19,20]. However, methods of how this information can be used still need investigation. In addition, it is important to balance the self-awareness of activity with perceived barriers that the participant may experience. Although using the activity tracker helped keep the goal of increasing activity at the top of participants' minds, but for most, the tracker was not a sufficient motivator to overcome personal barriers and achieve a significant increase in activity (change in routine). During closeout interviews, we were able to separate participants by those who seemed to be motivated by extrinsic versus intrinsic motivators. We found that those who expressed an extrinsic motivator to pursue an activity were more influenced by wearing their activity tracker, whereas those intrinsically motivated to pursue an activity were less likely to express being affected by tracker goals. These participants showed more commitment to making changes to their routine and increasing their activity level, albeit small changes. The fourth poignant point was if the participant did not already have another trigger of activity, such as working with a trainer or other personal fitness/activity goal that worked for the participant, the tracker was more likely to trigger a change in activity level for the participant. Although not explored during this study, previous research has indicated that if a person is not meeting his or her activity goals, the use of a tracker can be discouraging [21]. It would be interesting to pursue this in a follow-up study to determine if a combination of how people are motivated and the feedback they receive on achieving their goals affects overall engagement with trackers. Although this study was short, 9 of the 30 enrolled participants dropped out before the end of the study period. We hypothesize that as participants were given the activity tracker in the beginning of the study and were permitted to keep the device, participants were not as motivated to remain in the study and complete closeout surveys without further recompense. We have modified further phases of this study to deliver compensation at additional timepoints to attain a more complete data set. As an extension to this study, we plan to develop a machine learning–based app to encourage tracker users to stay engaged with exercise. We will look to include questions to elucidate the type of information that could assist in determining methods for keeping participants engaged dependent on whether they are driven by more intrinsic versus extrinsic motivators. This study had a few limitations. Target enrollment was low, and all participants were recruited from the same lab, which may limit the generalizability of the study. In addition, a larger sample population may have resulted in more significant results. Socioeconomic status (SES) was not collected from participants;

however, education, which can affect SES, was used as a proxy with no difference between engagement group and education reported. Finally, participants choosing to be in this study may also be more likely to engage with a FitBit tracker compared with those not choosing to participate in the study. Overall, as part of this study, we were able to gain many insights into why overweight participants may or may not engage with their activity tracker. This information will be used to create an algorithm to better sustain engagement with activity trackers.

Acknowledgments

This research was supported in conjunction with a grant from the Robert Wood Johnson Foundation (grant identification number: 71963).

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. J Appl Physiol (1985) 2005 Sep;99(3):1193-1204. [doi: 10.1152/japplphysiol.00160.2005] [Medline: 16103522]
2. Carlson SA, Fulton JE, Pratt M, Yang Z, Adams EK. Inadequate physical activity and health care expenditures in the United States. Prog Cardiovasc Dis 2015;57(4):315-323 [FREE Full text] [doi: 10.1016/j.pcad.2014.08.002] [Medline: 25559060]
3. 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 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
4. Sparling PB, Howard BJ, Dunstan DW, Owen N. Recommendations for physical activity in older adults. Br Med J 2015 Jan 21;350:h100. [doi: 10.1136/bmj.h100] [Medline: 25608694]
5. Eheman C, Henley SJ, Ballard-Barbash R, Jacobs EJ, Schymura MJ, Noone AM, 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 May 1;118(9):2338-2366 [FREE Full text] [doi: 10.1002/cncr.27514] [Medline: 22460733]
6. Ryerson AB, Eheman CR, Altekruse SF, Ward JW, Jemal A, Sherman RL, et al. Annual report to the nation on the status of cancer, 1975-2012, featuring the increasing incidence of liver cancer. Cancer 2016 Dec 1;122(9):1312-1337 [FREE Full text] [doi: 10.1002/cncr.29936] [Medline: 26959385]
7. Fox S, Duggan M. Pew Research Center. 2013. Tracking for Health URL: http://www.pewinternet.org/2013/01/28/tracking-for-health/
8. 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. J Mob Technol Med 2016 Jul;5(2):2-11 [FREE Full text] [doi: 10.7309/jmtm.5.2.2] [Medline: 27493694]
9. Fox G, Garland S, Keibel A, Saxon LA. NEJM Catalyst - Practical Innovations in Health Care Delivery. 2017. Why People Stick With or Abandon Wearable Devices URL: http://catalyst.nejm.org/stay-abandon-wearable-devices/?utm_campaign=Connect
10. Södlerlund 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;129(3):132-142. [doi: 10.1177/1757913908094805] [Medline: 19514637]
11. 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 (Silver Spring) 2011 Jan;19(1):100-109 [FREE Full text] [doi: 10.1038/oby.2010.122] [Medline: 20539299]
12. Bickmore T, Schulman D, Yin L. Maintaining engagement in long-term interventions with relational agents. Appl Artif Intell 2010 Jul 1;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
13. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exerc Psychol 2004 Jun;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
14. Centers for Disease Control and Prevention. 2007. Barriers to Physical Activity URL: https://www.cdc.gov/cancer/dcpc/prevention/policies_practices/physical_activity/barriers.htm
15. Holmen H, Wahl A, Torbjørnsen A, Jenum A, 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 Diabetes Res Care 2016;4(1):e000193 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000193] [Medline: 27239317]
16. 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 Sep;18(7):873-880 [FREE Full text] [doi: 10.1007/s11136-009-9496-9] [Medline: 19543809]

17. 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. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
18. Rosenberg D, Kadokura EA, Bouldin ED, Miyawaki CE, Higano CS, Hartzler AL. Acceptability of FitBit for physical activity tracking within clinical care among men with prostate cancer. AMIA Annu Symp Proc 2016;2016:1050-1059 [FREE Full text] [Medline: 28269902]
19. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health 2017 Dec 15;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]
20. Fritz T, Huang EM, Murphy GC, Zimmermann T. Persuasive Technology in the Real World: A Study of Long-Term Use of Activity Sensing Devices for Fitness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI'14; April 26-May 1, 2014; Toronto, Ontario, Canada p. 487-496. [doi: 10.1145/2556288.2557383]
21. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on FitBits with adolescents. Am J Health Educ 2017 Jul 28;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]

Abbreviations

BBA: Barriers to Being Active
BREQ-2: Behavioral Regulation in Exercise Questionnaire
GPH: global physical health
PHQ: Patient Health Questionnaire
PROMIS: Patient-Reported Outcomes Measurement Information System
SES: socioeconomic status

Edited by G Eysenbach; submitted 16.07.18; peer-reviewed by C Burns, D Bouchard, S Edney, H Yang; comments to author 17.08.18; revised version received 11.01.19; accepted 11.02.19; published 24.10.19.

Please cite as:
Centi AJ, Atif M, Golas SB, Mohammadi R, Kamarthi S, Agboola S, Kvedar JC, Jethwani K
Factors Influencing Exercise Engagement When Using Activity Trackers: Nonrandomized Pilot Study
JMIR Mhealth Uhealth 2019;7(10):e11603
URL: https://mhealth.jmir.org/2019/10/e11603
doi:10.2196/11603
PMID:31651405

Original Paper

A Mobile Health App (Roadmap 2.0) for Patients Undergoing Hematopoietic Stem Cell Transplant: Qualitative Study on Family Caregivers' Perspectives and Design Considerations

Dima Chaar[1], BA, MHI; Ji Youn Shin[2], BFA, MDes; Amanda Mazzoli[3], BSc; Rebecca Vue[3]; Jacob Kedroske[3]; Grant Chappell[3]; David A Hanauer[4], MS, MD; Debra Barton[5], PhD; Afton L Hassett[6], PhD; Sung Won Choi[3], MS, MD

[1]University of Michigan, School of Public Health, Department of Epidemiology, Ann Arbor, MI, United States

[2]Michigan State University, College of Communication Arts and Sciences, Department of Media and Information, East Lansing, MI, United States

[3]University of Michigan, Medical School, Blood and Marrow Transplantation Program, Ann Arbor, MI, United States

[4]University of Michigan, Medical School, Michigan Institute for Clinical and Health Research, Ann Arbor, MI, United States

[5]University of Michigan, School of Nursing, Ann Arbor, MI, United States

[6]University of Michigan, Medical School, Department of Anesthesiology, Ann Arbor, MI, United States

Corresponding Author:
Sung Won Choi, MS, MD
University of Michigan
Medical School
Blood and Marrow Transplantation Program
Medical Professional Building, D4118 SPC 5718
1500 E Medical Center Dr
Ann Arbor, MI, 48109
United States
Phone: 1 734 615 5707
Email: sungchoi@umich.edu

Abstract

Background: Hematopoietic stem cell transplantation (HCT), also referred to as blood and marrow transplantation (BMT), is a high-risk, but potentially curative therapy for a number of cancer and noncancer conditions. BMT Roadmap (Roadmap 1.0) is a mobile health app that was developed as a family caregiver–facing tool to provide informational needs about the health status of patients undergoing inpatient HCT.

Objective: This study explored the views and perceptions of family caregivers of patients undergoing HCT and their input regarding further technology development and expansion of BMT Roadmap into the outpatient setting (referred to as Roadmap 2.0).

Methods: Semistructured qualitative interviews were conducted among 24 family caregivers. Questions were developed from existing literature coupled with prior in-depth observations and interviews in hospital-based settings to explore the study objectives. Participants were recruited during routine outpatient clinic appointments of HCT patients, and all interviews were conducted in the participants' homes, the setting in which Roadmap 2.0 is intended for use. A thematic analysis was performed using a consistent set of codes derived from our prior research. New emerging codes were also included, and the coding structure was refined with iterative cycles of coding and data collection.

Results: Four major themes emerged through our qualitative analysis: (1) stress related to balancing caregiving duties; (2) learning and adapting to new routines (resilience); (3) balancing one's own needs with the patient's needs (insight); and (4) benefits of caregiving. When caregivers were further probed about their views on engagement with positive activity interventions (ie, pleasant activities that promote positive emotions and well-being such as expressing gratitude or engaging in activities that promote positive thoughts, emotions, and behaviors), they preferred a "menu" of positive activities to help support caregiver health and well-being.

Conclusions: This study involved family caregivers as participants in the development of new components for Roadmap 2.0. Our research provided a further understanding of the many priorities that hematopoietic stem cell transplant family caregivers face while maintaining balance in their lives. Their schedules can often be unpredictable, even more so once the patient is

discharged from the hospital. Our findings suggest that expanding Roadmap 2.0 into the outpatient setting may provide critical caregiver support and that HCT caregivers are interested in and willing to engage in positive activities that may enhance well-being and attenuate the stress associated with caregiving.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.4918

(JMIR Mhealth Uhealth 2019;7(10):e15775) doi:10.2196/15775

KEYWORDS

caregivers; allogeneic hematopoietic stem cell transplant; home interviews; user experience; mobile health apps

Introduction

Family Caregivers in the Clinical Context of Hematopoietic Stem Cell Transplantation

Hematopoietic stem cell transplantation, also commonly referred to as blood and marrow transplantation (BMT), is a high-risk, but potentially curative therapy for a number of cancer and noncancer conditions [1]. Given the high-risk associated with this treatment modality, patients require a dedicated family (or friend) caregiver who is available full time (24/7) for at least the first 100 days after the transplant [2]. The caregiver must receive medical training on catheter line care, intravenous and oral medication management, identification of signs and symptoms of infection or other posttransplant complications, and management of care and activities of daily living. Significant levels of anxiety and distress are experienced by caregivers during the peritransplant period (ie, pretransplant and immediately following discharge from the hospital for the hematopoietic stem cell transplant procedure) [3]. The stress and demands of caregiving for HCT patients may be further heightened if patients and caregivers are required to relocate to be closer to the hospital. This burden is associated with a significant decline in emotional, physical, and social functioning of the caregiver [4-7]. The reciprocal relationship between patient and caregiver emotional distress strongly supports the need to promote caregiver health and well-being [8].

Need for Novel Caregiver-Based Interventions

Receiving education on patient illness has been identified as an approach for reducing stress and anxiety [9,10]. Complex medical tasks are initially performed by a professional nurse in the hospital setting, with responsibility for these tasks transferred to the caregiver to perform at home. Accordingly, the caregiver needs education about the diagnosis and prognosis of the patient (knowledge), training in medical care (skills), and information about available resources to achieve patient and caregiver-centered care [11], which may protect the caregiver from harm by increasing their health and well-being [12]. Thus, novel interventions that provide caregiver support are urgently needed [13].

Mobile Health Technology: BMT Roadmap

BMT Roadmap (Roadmap 1.0) was developed as a caregiver-facing mobile health app to provide information, education, and skills building to meet informational needs during the inpatient transplant process (Multimedia Appendix 1) [14-19]. Screenshots of the app are provided in Multimedia Appendix 2. Patients, caregivers, and health care providers actively participated in iterative cycles of user-centered design, development, and testing, which has been shown to be crucial for the correct use and adoption [20]. To date, >100 HCT caregivers and patients have consented and enrolled in institutional review board–approved studies to assess the feasibility of implementing Roadmap 1.0 during the inpatient setting [21]. Patient-reported outcomes were collected at baseline, discharge, and day 100 in a pre- and poststudy design.

Roadmap 1.0 was designed as a multicomponent app that included the following modules: an overview of the criteria needed for discharge along with training for self-management, which was required for discharge (eg, videos of central line care); real-time laboratory results graphed so that visual trends could be tracked longitudinally across time; personalized medication lists with common indications, dosing and schedule, and side effects; health care provider “Facebook” photos; details about clinical trials that patients were enrolled in; and educational materials defining commonly used medical terms and concepts that the medical team discusses with families. Caregivers were instructed to use it freely throughout their hospital stay.

Roadmap 1.0 was rated as highly useful and easy to use [22,23]; caregiver anxiety [24] decreased; caregiver health-related quality of life [25] improved; and caregiver activation [26] increased at discharge compared with baseline [22]. The most used modules were laboratory studies and medications, which were rated highly useful and easy to understand and visualize. In addition, weekly semistructured qualitative interviews were conducted on topics including acceptance of and barriers to Roadmap 1.0, informational needs, and psychosocial burdens. Major themes emerging from content analyses using memos and coding included the following: Roadmap 1.0 had high usefulness, ease of use, and likeability; its laboratory module was the most viewed; and participants expressed a desire for additional components along with expansion of Roadmap 1.0 into the outpatient setting. The interviews also highlighted the need for supporting caregiver health and well-being, but also showed the positive aspects of caregiving, such as new or closer relationships with loved ones. Collectively, these findings support the rationale to extend Roadmap to the outpatient setting. Furthermore, caregivers expressed modules that include caregiver-specific resources as well as positive activities that they could perform throughout the transplant trajectory, since it does not end after discharge from the hospital.

This study aimed to explore the views and perceptions of HCT caregivers for expanding Roadmap 1.0 into the home environment, outside of the hospital, where the app would be

most used (the expanded version of the app will be referred to as Roadmap 2.0 henceforth). We were also interested in examining the life of a caregiver in her/his home environment to inform Roadmap 2.0.

Methods

Study Site

The field sites in this study included participant's homes, defined as permanent residence, rental apartment, and local extended-stay accommodations (eg, motels, hotels, and suites). The homes were all within approximately 90 minutes of travel distance from the University of Michigan, a large, tertiary academic medical center, as mandated by the transplant program.

Study Design, Recruitment, and Consent

Semistructured interviews were conducted among 24 family caregivers. Eligibility for study participation included the following: primary family caregiver who had already experienced the transplant procedure with their loved one (patient) and was in the posttransplant phase of care; age≥18 years; comfortable with reading and speaking English; willing to participate in a face-to-face interview in a home setting; and able to provide informed consent. Participants were recruited through referrals from the hematopoietic stem cell transplant clinical team and included caregivers of transplant patients visiting the hospital for routine outpatient clinic appointments. The hematopoietic stem cell transplant clinical team notified the research team if the caregiver agreed to learn more about the study. All caregivers approached for the study agreed to participate. One caregiver was approached by the study coordinator, but was deemed ineligible because English was not her primary language and she was not comfortable with reading and speaking English. Our primary dataset included semistructured interviews with the primary caregivers of hematopoietic stem cell transplant patients (ie, provided ≥50% of caregiving responsibilities). Saturation is defined in qualitative research as a criterion for discontinuing data collection and/or analysis [27]. Recruitment ended once it was determined that no additional data were being found, through which the investigators could develop new thematic categories.

Ethical approval for this study was obtained by the Institutional Review Board. All participants signed an informed consent prior to the research investigators traveling to their residence. Once the informed consent document was signed, the research team worked with the participant to arrange for a date and time that worked well for both. All interviews were conducted in the participants' homes. A minimum of two research assistants traveled to the homes. The home semistructured interview questions were informed by existing literature as well as prior observations and interviews conducted in hospital-based settings (inpatient and outpatient transplant units). The questions were subsequently optimized by the study team. Approximately 250 hours of hospital-based observations and interviews were conducted by a minimum of two research assistants (SW Choi, unpublished data, 2019). Observations were recorded as field notes, and interviews were audio-recorded, with permission, and subsequently professionally transcribed (Babbletype LLC, Philadelphia, Pennsylvania). The home interviews explored several domains including demographic information, general caregiver duties after transplant, posttransplant life experiences, use of technology and social media platforms to support caregiver self-management and provide informational resources, and positive activity interventions to promote caregiver health and well-being (Multimedia Appendix 3).

Positive Activity Interventions

Based on prior in-depth observations and interviews conducted in the hospital of Roadmap 1.0, we found that caregivers were interested in and willing to engage in positive activity interventions. It was acknowledged that caregiving can be stressful and personal well-being needed to be attended to in order to best support the patient. Thus, in this study, positive activity interventions were defined to caregivers as simple pleasant activities that were intentional and could be developed into routine practices such as expressing gratitude or scheduling activities that promoted positive thoughts, emotions, or behaviors [28-30]. Artifacts, such as photographs, that reflected the caregiver home environment were also collected during home interviews in efforts to inform Roadmap 2.0.

Participant Demographics

The median age of the study participants was 57 years (range, 31-71 years). The majority were married or in a domestic partnership (23/24 [96%]), white (23/24 [96%]), and female (20/24 [83%]). In addition, 20/24 (83%) received at least a 2-year college degree and 12 participants (50%) received at least a 4-year college degree. The interviews took place in permanent residences (11/24 [46%]), rental apartments (10/24 [42%]), or local accommodations (3/24 [12%]; Multimedia Appendix 4).

Thematic Analysis

The semistructured interviews were conducted one-to-one by research investigators who were trained in qualitative methods with experience working in the hematopoietic stem cell transplant population. The interviews lasted about 60-80 minutes. The interview script was developed by DC and JS, and minor iterations were made to adjust for clarity of the questions after piloting the home interview script with the rest of the research team (SC, DH, DC, JK, RV, and GC; Multimedia Appendix 5). The sentence structures were refined to make the question as well as potential probing questions clear. All qualitative semistructured interviews were audio-recorded, deidentified, and entered into NVivo Pro 11 (QSR International, Melbourne, Australia). Data were analyzed using a thematic approach, in that our data collection and analyses mutually informed one another. During the first stage of data analysis, initial interviews were professionally transcribed verbatim, followed by qualitative content analysis [31] of the data using an open coding method. The transcripts were coded according to each concept (ie, line by line or by paragraph) using NVivo Pro 11 software. The research team (SC, DC, JS, JK, RV, and GC) then collectively generated new codes as significant concepts, and patterns of data (themes) were identified, discussed, and revised. All the members who participated in the thematic analysis followed the phases of thematic analysis:

(1) familiarized her/himself with the data; (2) generated initial codes; (3) searched for themes; (4) reviewed themes; (5) defined and named the themes; and (6) produced the report [32].

The second stage of data analysis, using new interview as well as artifact data that emerged, resulted in consistent themes and confirmed our findings. A minimum of three team members took an active role in identifying patterns/themes based on how the data were classified by codes. Common structures and themes that arose repeatedly through content analysis were identified and organized hierarchically by patterns or structures in the code [33], which were subsequently further grouped together in overarching themes. The final interpretation of codes and themes were completed by the first and second authors and two senior researchers (DC, JS, DH, SC). The semistructured interview script and code book are provided in Multimedia Appendices 5 and 6, respectively. Texts or artifacts were gathered from participant homes and analyzed with the qualitative data [34] to further inform the themes that were emerging.

Results

Overview

Four overarching themes emerged through 37 codes that developed out of our thematic qualitative analyses: (1) stress related to balancing caregiving duties, (2) learning and adapting to new routines (resilience), (3) balancing one's own needs with the patient's needs (insight), and (4) benefits of caregiving.

Theme 1: Stress Related to Balancing Caregiving Duties

Once patients were discharged home, caregivers reported heightened anxiety of the new duties and tasks as well as being in the small living conditions of temporary housing (eg, rental apartment or hotel room; Multimedia Appendix 7). After a prolonged hospital course, caregivers described further time away from their (permanent resident) homes and support system as being isolating. Caregivers were suddenly thrust into multitasking and balancing caregiving duties such as cleaning, providing transportation, planning meals, scheduling appointments, administering/organizing medications, and performing medical tasks (eg, central line care and infusions). Representative quotes are shown in Multimedia Appendix 8.

Theme 2: Learning and Adapting to New Routines (Resilience)

Caregivers reported that while they initially felt overwhelmed when they first came home after a prolonged transplant course in the hospital, they adjusted to the new routine of home care. Caregivers reported their ability to perform complex and demanding tasks that gradually improved over time through repeating the skill (eg, medication administration and dressing changes). A representative participant quote is shown in Multimedia Appendix 8. Caregivers developed their own processes for organizing medications according to time of administration (eg, morning, afternoon, and evening) (Multimedia Appendix 7); caregivers scheduled reminders in calendars (on paper) or used mobile devices to set alarms; caregivers saved business cards of various health care providers and taped them to their walls (Multimedia Appendix 7); caregivers also posted discharge contact information as well as local area dining and shopping options (Multimedia Appendix 7).

Despite the burden of these new duties and routines, some caregivers voiced optimism and strength. Having beautiful flowers or Get Well cards in the room provided encouragement and support (Multimedia Appendix 7). For example, despite the medically complex and emotionally difficult treatment process endured by the patient, a caregiver described it as "temporary" and the cancer diagnosis being a "bump in the road." Caregivers displayed resiliency in the midst of the transplant journey as they learned to adapt to the requirements of caregiving and shared new experiences with family or friends (Multimedia Appendix 8).

Theme 3: Balancing One's Own Needs With the Patient's Needs (Insight)

Caregivers gained insight by recognizing the need to care for their own health while also serving their loved ones. In addition to developing new routines on behalf of the patient, many were also adjusting to living in new, temporary accommodations in order to be closer to the hospital, as mandated by the transplant team. More than half of the study participants (52%) were not living in their permanent residence. As such, caregivers developed new strategies over the transplant trajectory that focused on themselves as well as the patients. Many of these strategies included self-care or management of their own stress that enabled them to be "better" caregivers. Some examples that were provided included taking a quiet moment for themselves (ie, quiet time or moments of solitude), focusing on sleep, maintaining pre-existing relationships or friendships, and engaging in sports or physical activity (eg, golf and yoga). One participant noted that simply getting out of the house was useful (Multimedia Appendix 8).

Theme 4: Benefits of Caregiving

Caregivers made considerable adjustments in their daily lives. Some caregivers took time off work and had not yet returned to work at the time of interviews. Nonetheless, caregivers found meaning in performing their medically related duties on behalf of the patient or simply from spending time with the patient in "isolated" environments (ie, minimizing visitors or going to the grocery store during early morning or later evening hours to mitigate infectious contacts). Caregivers noticed that they found benefits in caregiving, such as deeper compassion and empathy toward the patient; they identified the positive aspects of caregiving. For example, while one caregiver shared that she now had to rely on other family members, this allowed deeper relationships to form (Multimedia Appendix 8). Another caregiver expressed how journaling provided an opportunity to reflect on the transplant journey (Multimedia Appendix 7).

Positive Activity Interventions

Caregivers were asked to rank a particular positive activity on a scale of 1-10, with the higher number indicating that the activity could be very beneficial (Multimedia Appendix 3). Pleasant Activity Scheduling, where participants would set aside

time each day for a positive activity (eg, having ice cream with a friend or taking a walk in the park) was ranked the highest. This activity was followed by Gratitude journaling (ie, writing down things for which they were grateful) and Savoring the Moment (ie, being mindful of and savoring a pleasant experience, such as drinking coffee or spending time with a friend, and using all of the senses to solidify the memory; Table 1). When reflecting on the possible benefits of the Savoring the Moment positive activity, one of the caregivers shared how he was already incorporating this activity to his early mornings and found it helpful (Multimedia Appendix 8).

Table 1. Engagement with positive activity interventions.

Positivity activity	Median Likert rating (range)[a]
Pleasant Activity Scheduling	9.0 (5.0-10.0)
Gratitude Journaling	10.0 (3.0-10.0)
Savoring	9.0 (4.0-10.0)
Random Acts of Kindness	8.0 (2.0-10.0)
Positive Piggy Bank	6.0 (1.0-10.0)
Signature Strengths	6.5 (1.0-9.0)

[a]Scale: 1-10; 1=not likely to engage in the activity; 10=very likely to engage in the activity.

Discussion

Main Findings

In this study, we explored the views and perceptions of HCT caregivers to expand Roadmap 2.0 into the home environment. Through our qualitative findings, four major themes emerged: (1) stress related to balancing caregiving duties; (2) learning and adapting to new routines (resilience); (3) balancing one's own needs with the patient's needs (insight); and (4) benefits of caregiving. These findings support the importance of linking caregivers to resources throughout the transplant trajectory, beyond the hospital course, because in many instances, caregivers are not even aware of the support services available. The artifacts collected (eg, photographs) and semistructured interviews in caregivers' home environment provided content and design considerations for the outpatient version of Roadmap 2.0; suggested topics or categories are provided in Multimedia Appendix 3.

Caregivers in this study gained new insights and recognized the importance of maintaining their own health and well-being, while also caring for the patient. Accordingly, they indicated the willingness to engage in positive activity exercises. The research team asked the caregivers to rank each activity on a Likert scale. Caregivers reflected on their own subjective values and experiences in determining the ranking. In general, they preferred having a "menu" or choice of a variety of positive activity exercises. The majority favored Pleasant Activity Scheduling, where caregivers would set aside time each day for a positive activity. This was followed by Gratitude journaling and Savoring the Moment, a mindfulness-based exercise. Interestingly, these activities correlated with the examples caregivers provided when discussing strategies employed while caring for the patient (ie, setting aside time for themselves).

Positive Aspects of Caregiving

Increasing amounts of data suggest that simple strategies aimed at enhancing positive thoughts, emotions, and behaviors are effective and highly scalable [28-30]. Positive activity interventions such as daily positive reflection, gratitude journals, and conducting acts of kindness have been used in other medical populations including those with cardiovascular disease, cancer, diabetes, and chronic pain [35-38], but not in caregivers of HCT patients. Our preliminary data suggest that caregivers desire the opportunity to try these activities to reduce stress and improve their health-related quality of life. Research has shown that much focus has been placed on the wide range of negative implications associated with caregiving [4], such as increases in depression, anxiety, and other health-related concerns [39]. Despite this, a majority of caregivers have recognized the benefits of caregiving [40,41].

Studies have shown that positive psychology interventions enhance subjective and psychological well-being and help reduce depressive symptoms in a cost-effective and targeted manner, particularly when combined with other evidence-based interventions [29]. Further, research supports implementing a selection of positive activity exercises as opposed to introducing only one activity (ie, Gratitude journal, Random Acts of Kindness, and Signature Strengths vs Gratitude journal only). Our caregiver interviews herein highlight the importance that "one size does not fit all," despite the relatively homogenous participant population. It is possible that caregivers who are able to choose their positive activity exercise may want to engage in them longer and frequently. Nonetheless, research indicates that the ideal dosage and timing as well as combination of activities remain unclear [42].

Home Environment as a Novel Study Site to Explore Caregivers' Views

HCT caregiving is intense and complex with unexpected fluctuations throughout the transplant trajectory. This paper has specifically focused on descriptive qualitative approaches conducted in the home setting of HCT caregivers. We focused on the home environment, which was informed from prior observational and qualitative work in the hospital. Since the early prototype of Roadmap 1.0 [43,44], we have continued to incorporate key stakeholders (eg, patients, caregivers, physicians, nurses, psychologists, and social workers) in its design and development [43,44]. Importantly, the iterative

process has included a multidisciplinary team. In the home environment, it was apparent that HCT caregivers face competing demands related to caregiving while maintaining other life duties (ie, self-care, home care, work, and other children, spouse, or parent care). More strikingly, even after hospital discharge, more than half our HCT caregivers were living in rental apartments or local accommodations. The unfamiliar and isolated environment critically speaks to novel and innovative solutions that are needed to support them beyond the hospital setting. We need to design, develop, and test interventions that support caregiver health and well-being in the midst of multitasking and balancing duties while further strengthening their resiliency. Mobile health technology has the potential to deliver scalable solutions [45-48]. Although HCT caregivers sought out new or different relationships with family members or friends, providing (reputable or health care provider-approved) caregiver-specific resources may provide benefit.

Limitations and Future Directions

Our work supports caregivers as key stakeholders in the development of mobile health technology for the provision of patient care. Although we recognize the limitations of this study, such as single-center design, and relatively homogeneous participant population (eg, white, female, educated) with bias toward willingness to participate, the strengths of the study include rigorous data collection and analyses and novel study site (home environment). Based on our findings, we have constructed a Caregiver Health Survey to quantify the views and perspectives of HCT caregivers nationally, which will capture a larger, more diverse population. We hope that our collective qualitative and quantitative findings will inform the design and development of a positive psychology intervention to provide caregiver support in the outpatient setting. The multicomponent intervention will include a “menu” of positive activities that will be tested in a randomized controlled trial design.

Acknowledgments

This work was supported by an American Society of Hematology Bridge Grant and NIH/NHBLI grant (1R01HL146354) and the Edith S Briskin and Shirley K Schlafer Foundation (author SWC). We wish to thank the caregivers and patients who participated in this study.

Authors' Contributions

DC: writing of the original draft, data curation, data analysis, and visualization; JYS: data curation, data interpretation, and writing-review/editing; AM: patient recruitment, data collection, data curation, data interpretation, and writing-review/editing; RV: data collection, data curation, data analysis, visualization, and writing-review/editing; JK: data collection, data curation, data analysis, and writing-review/editing; GC: data collection, data curation, data analysis, and writing-review/editing; DH: data curation, visualization, and writing-review/editing; DB: data interpretation, and writing-review/editing; AH: data interpretation and writing-review/editing; SWC: data curation, investigation, methodology, data analysis, resources, supervision, visualization, writing-original draft, and writing-review/editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
BMT Roadmap components.
[PDF File (Adobe PDF File), 74 KB - mhealth_v7i10e15775_app1.pdf]

Multimedia Appendix 2
Screenshots of BMT Roadmap laboratory studies and medications modules.
[PDF File (Adobe PDF File), 211 KB - mhealth_v7i10e15775_app2.pdf]

Multimedia Appendix 3
Caregiver-specific resources and positive activities components.
[PDF File (Adobe PDF File), 79 KB - mhealth_v7i10e15775_app3.pdf]

Multimedia Appendix 4
Participant demographics.
[PDF File (Adobe PDF File), 139 KB - mhealth_v7i10e15775_app4.pdf]

Multimedia Appendix 5
Home interview script.
[PDF File (Adobe PDF File), 137 KB - mhealth_v7i10e15775_app5.pdf]

Multimedia Appendix 6
Codebook.
[PDF File (Adobe PDF File), 139 KB - mhealth_v7i10e15775_app6.pdf]

Multimedia Appendix 7
Artifacts collected in the home environment.
[PDF File (Adobe PDF File), 826 KB - mhealth_v7i10e15775_app7.pdf]

Multimedia Appendix 8
Participant quotes.
[PDF File (Adobe PDF File), 86 KB - mhealth_v7i10e15775_app8.pdf]

References

1. Copelan EA. Hematopoietic stem-cell transplantation. N Engl J Med 2006 Apr 27;354(17):1813-1826. [doi: 10.1056/NEJMra052638] [Medline: 16641398]
2. Laudenslager ML, Simoneau TL, Kilbourn K, Natvig C, Philips S, Spradley J, et al. A randomized control trial of a psychosocial intervention for caregivers of allogeneic hematopoietic stem cell transplant patients: effects on distress. Bone Marrow Transplant 2015 Aug;50(8):1110-1118 [FREE Full text] [doi: 10.1038/bmt.2015.104] [Medline: 25961767]
3. Simoneau TL, Mikulich-Gilbertson SK, Natvig C, Kilbourn K, Spradley J, Grzywa-Cobb R, et al. Elevated peri-transplant distress in caregivers of allogeneic blood or marrow transplant patients. Psychooncology 2013 Sep;22(9):2064-2070 [FREE Full text] [doi: 10.1002/pon.3259] [Medline: 23440998]
4. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003 Jun;18(2):250-267. [Medline: 12825775]
5. Vitaliano PP, Zhang J, Scanlan JM. Is caregiving hazardous to one's physical health? A meta-analysis. Psychol Bull 2003 Nov;129(6):946-972. [doi: 10.1037/0033-2909.129.6.946] [Medline: 14599289]
6. Stenberg U, Ruland CM, Miaskowski C. Review of the literature on the effects of caring for a patient with cancer. Psychooncology 2010 Oct;19(10):1013-1025. [doi: 10.1002/pon.1670] [Medline: 20014159]
7. Bevans M, Sternberg EM. Caregiving burden, stress, and health effects among family caregivers of adult cancer patients. JAMA 2012 Jan 25;307(4):398-403 [FREE Full text] [doi: 10.1001/jama.2012.29] [Medline: 22274687]
8. Northouse LL, Katapodi MC, Schafenacker AM, Weiss D. The impact of caregiving on the psychological well-being of family caregivers and cancer patients. Semin Oncol Nurs 2012 Nov;28(4):236-245 [FREE Full text] [doi: 10.1016/j.soncn.2012.09.006] [Medline: 23107181]
9. Vaughan C, Trail TE, Mahmud A, Dellva S, Tanielian T, Friedman E. Informal Caregivers' Experiences and Perceptions of a Web-Based Peer Support Network: Mixed-Methods Study. J Med Internet Res 2018 Aug 28;20(8):e257 [FREE Full text] [doi: 10.2196/jmir.9895] [Medline: 30154074]
10. Latulippe K, Tremblay M, Poulin V, Provencher V, Giguere AM, Sévigny A, et al. Prioritizing the Needs of Caregivers of Older Adults to Support Their Help-Seeking Process as a First Step to Developing an eHealth Tool: The Technique for Research of Information by Animation of a Group of Experts (TRIAGE) Method. JMIR Aging 2019 May 23;2(1):e12271 [FREE Full text] [doi: 10.2196/12271] [Medline: 31518269]
11. Gillick MR. The critical role of caregivers in achieving patient-centered care. JAMA 2013 Aug 14;310(6):575-576. [doi: 10.1001/jama.2013.7310] [Medline: 23867885]
12. Reinhard S, Given B, Petlick N, Bemis A. Chapter 14 Supporting Family Caregivers in Providing Care. In: Hughes RG, editor. Patient Safety Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality; Apr 2008.
13. Milosevic M, Singh P. The Informal Caregiver Engagement Framework. iproc 2017 Sep 22;3(1):e8. [doi: 10.2196/iproc.8374]
14. Choi SW, Chang L, Hanauer DA, Shaffer-Hartman J, Teitelbaum D, Lewis I, et al. Rapid reduction of central line infections in hospitalized pediatric oncology patients through simple quality improvement methods. Pediatr. Blood Cancer 2012 Apr 22;60(2):262-269. [doi: 10.1002/pbc.24187] [Medline: 22522576]
15. Büyüktür AG, Ackerman MS. Issues and opportunities in transitions from speciality care: a field study of bone marrow transplant. Behaviour & Information Technology 2014 Jun 09;34(6):566-584. [doi: 10.1080/0144929x.2014.921729]
16. Keusch F, Rao R, Chang L, Lepkowski J, Reddy P, Choi SW. Participation in clinical research: perspectives of adult patients and parents of pediatric patients undergoing hematopoietic stem cell transplantation. Biol Blood Marrow Transplant 2014 Oct;20(10):1604-1611 [FREE Full text] [doi: 10.1016/j.bbmt.2014.06.020] [Medline: 24972252]
17. Kaziunas E, Buyuktur AG, Jones J, Choi SW, Hanauer DA, Ackerman MS. Transition and Reflection in the Use of Health Information: The Case of Pediatric Bone Marrow Transplant Caregivers. 2015 Presented at: Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '15); 2015; Vancouver, British Columbia, Canada.

18. Kaziunas E, Hanauer DA, Ackerman MS, Choi SW. Identifying unmet informational needs in the inpatient setting to increase patient and caregiver engagement in the context of pediatric hematopoietic stem cell transplantation. J Am Med Inform Assoc 2016 Jan;23(1):94-104 [FREE Full text] [doi: 10.1093/jamia/ocv116] [Medline: 26510878]
19. Buyuktur A. University of Michigan, School of Information. 2015. Temporality and Information Work in Bone Marrow Transplant (PhD Dissertation Thesis) URL: https://deepblue.lib.umich.edu/handle/2027.42/111491
20. Cristancho-Lacroix V, Moulin F, Wrobel J, Batrancourt B, Plichart M, De Rotrou J, et al. A web-based program for informal caregivers of persons with Alzheimer's disease: an iterative user-centered design. JMIR Res Protoc 2014;3(3):e46 [FREE Full text] [doi: 10.2196/resprot.3607] [Medline: 25263541]
21. Grossman L, Choi S, Collins S, Dykes P, O'Leary K, Rizer M. Implementation of Acute Care Patient Portals: Recommendations on Utility and Use from Six Early Adopters. J Am Med Inform Assoc 2018 2018 Apr 01. [doi: 10.1093/jamia/ocx074]
22. Runaas L, Bischoff E, Hoodin F, Kentor R, Ostarello L, Seyedsalehi S. A Novel Health Informatics Tool to Improve Caregiver Activation: Findings from Pediatric BMT in a Hospital-Based Setting. Blood 2016;128:2382. [doi: 10.1182/blood.V128.22.2382.2382]
23. Fauer A, Runaas L, Bischoff E, Hoang T, Gupta R. User-centered health information technology tool for bloodmarrow transplant patients: A qualitative study. Washington, DC: AcademyHealth; 2017 Jun Presented at: Academy Health Annual Meeting; June 25-27, 2017; New Orleans, LA.
24. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.
25. Weitzner MA, Jacobsen PB, Wagner H, Friedland J, Cox C. The Caregiver Quality of Life Index-Cancer (CQOLC) scale: development and validation of an instrument to measure quality of life of the family caregiver of patients with cancer. Qual Life Res 1999;8(1-2):55-63. [Medline: 10457738]
26. Pennarola BW, Rodday AM, Mayer DK, Ratichek SJ, Davies SM, Syrjala KL, HSCT-CHESS Study. Factors associated with parental activation in pediatric hematopoietic stem cell transplant. Med Care Res Rev 2012 Apr;69(2):194-214 [FREE Full text] [doi: 10.1177/1077558711431460] [Medline: 22203645]
27. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
28. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
29. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
30. Hassett AL, Finan PH. The Role of Resilience in the Clinical Management of Chronic Pain. Curr Pain Headache Rep 2016 Jun;20(6):39. [doi: 10.1007/s11916-016-0567-7] [Medline: 27115770]
31. Krippendorff K. Content analysis: An introduction to its methodology. Newbury Park, CA: Sage Publications; 1980.
32. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Beyer H, Holtzblatt K. Contextual Design: Defining Customer-Centered Systems. San Francisco: Morgan Kaufmann Publishers, Inc; 1998.
34. Silverman D. Interpreting Qualitative Data, Second Edition. Thousand Oaks, CA: Sage Publications; 2001.
35. Huffman JC, Mastromauro CA, Boehm JK, Seabrook R, Fricchione GL, Denninger JW, et al. Development of a positive psychology intervention for patients with acute cardiovascular disease. Heart Int 2011 Sep 29;6(2):e14 [FREE Full text] [doi: 10.4081/hi.2011.e14] [Medline: 23825741]
36. Moskowitz JT, Hult JR, Duncan LG, Cohn MA, Maurer S, Bussolari C, et al. A positive affect intervention for people experiencing health-related stress: development and non-randomized pilot test. J Health Psychol 2012 Jul;17(5):676-692 [FREE Full text] [doi: 10.1177/1359105311425275] [Medline: 22021272]
37. Cohn MA, Pietrucha ME, Saslow LR, Hult JR, Moskowitz JT. An online positive affect skills intervention reduces depression in adults with type 2 diabetes. J Posit Psychol 2014 Jan 1;9(6):523-534 [FREE Full text] [doi: 10.1080/17439760.2014.920410] [Medline: 25214877]
38. Hausmann LRM, Parks A, Youk AO, Kwoh CK. Reduction of bodily pain in response to an online positive activities intervention. J Pain 2014 May;15(5):560-567. [doi: 10.1016/j.jpain.2014.02.004] [Medline: 24568751]
39. Barata A, Wood WA, Choi SW, Jim HSL. Unmet Needs for Psychosocial Care in Hematologic Malignancies and Hematopoietic Cell Transplant. Curr Hematol Malig Rep 2016 Aug;11(4):280-287. [doi: 10.1007/s11899-016-0328-z] [Medline: 27113094]
40. Farran CJ, Keane-Hagerty E, Salloway S, Kupferer S, Wilken CS. Finding Meaning: An Alternative Paradigm for Alzheimer's Disease Family Caregivers. The Gerontologist 1991 Aug 01;31(4):483-489. [doi: 10.1093/geront/31.4.483]
41. Sanders S. Is the glass half empty or full? Reflections on strain and gain in cargivers of individuals with Alzheimer's disease. Soc Work Health Care 2005;40(3):57-73. [doi: 10.1300/J010v40n03_04] [Medline: 15837668]

42. Parks AC, Della PMD, Pierce RS, Zilca R, Lyubomirsky S. Pursuing happiness in everyday life: the characteristics and behaviors of online happiness seekers. Emotion 2012 Dec;12(6):1222-1234. [doi: 10.1037/a0028587] [Medline: 22642345]
43. Maher M, Hanauer DA, Kaziunas E, Ackerman MS, Derry H, Forringer R, et al. A Novel Health Information Technology Communication System to Increase Caregiver Activation in the Context of Hospital-Based Pediatric Hematopoietic Cell Transplantation: A Pilot Study. JMIR Res Protoc 2015 Oct 27;4(4):e119 [FREE Full text] [doi: 10.2196/resprot.4918] [Medline: 26508379]
44. Maher M, Kaziunas E, Ackerman M, Derry H, Forringer R, Miller K, et al. User-Centered Design Groups to Engage Patients and Caregivers with a Personalized Health Information Technology Tool. Biol Blood Marrow Transplant 2016 Feb;22(2):349-358 [FREE Full text] [doi: 10.1016/j.bbmt.2015.08.032] [Medline: 26343948]
45. Shin JY, Kang TI, Noll RB, Choi SW. Supporting Caregivers of Patients With Cancer: A Summary of Technology-Mediated Interventions and Future Directions. Am Soc Clin Oncol Educ Book 2018 May 23(38):838-849 [FREE Full text] [doi: 10.1200/EDBK_201397] [Medline: 30231412]
46. Pensak NA, Joshi T, Simoneau T, Kilbourn K, Carr A, Kutner J, et al. Development of a Web-Based Intervention for Addressing Distress in Caregivers of Patients Receiving Stem Cell Transplants: Formative Evaluation With Qualitative Interviews and Focus Groups. JMIR Res Protoc 2017 Dec 22;6(6):e120 [FREE Full text] [doi: 10.2196/resprot.7075] [Medline: 28642213]
47. Shaffer KM, Chow PI, Cohn WF, Ingersoll KS, Ritterband LM. Informal Caregivers' Use of Internet-Based Health Resources: An Analysis of the Health Information National Trends Survey. JMIR Aging 2018 Dec 18;1(2):e11051 [FREE Full text] [doi: 10.2196/11051] [Medline: 31518244]
48. Turk A, Fairclough E, Grason Smith G, Lond B, Nanton V, Dale J. Exploring the Perceived Usefulness and Ease of Use of a Personalized Web-Based Resource (Care Companion) to Support Informal Caring: Qualitative Descriptive Study. JMIR Aging 2019 Aug 20;2(2):e13875 [FREE Full text] [doi: 10.2196/13875] [Medline: 31518272]

Abbreviations

BMT: blood and marrow transplantation
HCT: hematopoietic stem cell transplant

Edited by G Eysenbach; submitted 05.08.19; peer-reviewed by B Oh, A Mahnke, S Jalil, L Becker; comments to author 24.08.19; revised version received 30.08.19; accepted 23.09.19; published 24.10.19.

Please cite as:
Chaar D, Shin JY, Mazzoli A, Vue R, Kedroske J, Chappell G, Hanauer DA, Barton D, Hassett AL, Choi SW
A Mobile Health App (Roadmap 2.0) for Patients Undergoing Hematopoietic Stem Cell Transplant: Qualitative Study on Family Caregivers' Perspectives and Design Considerations
JMIR Mhealth Uhealth 2019;7(10):e15775
URL: http://mhealth.jmir.org/2019/10/e15775/
doi:10.2196/15775
PMID:31651402

Review

Assessment of the Efficacy, Safety, and Effectiveness of Weight Control and Obesity Management Mobile Health Interventions: Systematic Review

Elisa Puigdomenech Puig[1,2,3], MSc; Noemí Robles[2,3,4], PhD; Francesc Saigí-Rubió[5,6], PhD; Alberto Zamora[7,8], MD, PhD; Montse Moharra[1,9], BA; Guillermo Paluzie[7], MD, PhD; Mariona Balfegó[10], PhD; Guillem Cuatrecasas Cambra[10], MD; Pilar Garcia-Lorda[5], MD, PhD; Carme Carrion[2,3,4,5], PhD

[1]Agència de Qualitat i Avaluació Sanitàries de Catalunya, Barcelona, Spain

[2]Red de Investigación en Servicios de Salud en Enfermedades Crónicas, Barcelona, Spain

[3]eHealth Lab, Barcelona, Spain

[4]eHealth Center, Universitat Oberta de Catalunya, Barcelona, Spain

[5]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[6]Interdisciplinary Research Group on ICTs, Barcelona, Spain

[7]Corporació de Salut del Maresme i la Selva, Hospital de Blanes, Blanes, Spain

[8]Grup de Medicina Traslacional i Ciències de la Decisió, Departament de Ciències Mèdiques, Facultat de Medicina, Universitat de Girona, Girona, Spain

[9]CIBER Epidemiología y Salud Pública, Barcelona, Spain

[10]Clínica Sagrada Família, CPEN SL Servei d'Endocrinologia i Nutrició, Barcelona, Spain

Corresponding Author:
Carme Carrion, PhD
Faculty of Health Sciences
Universitat Oberta de Catalunya
Rambla del Poblenou, 156
Barcelona
Spain
Phone: 34 934 505 254
Email: mcarrionr@uoc.edu

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2023/1/e47585

Abstract

Background: The use of apps to tackle overweight and obesity by tracking physical and dietary patterns and providing recommendations and motivation strategies to achieve personalized goals has increased over recent years. However, evidence of the efficacy, effectiveness, and safety of these apps is severely lacking.

Objective: The aim of this study was to identify efficacy, safety, and effectiveness criteria used to assess weight control, overweight, and obesity management in mobile health (mHealth) interventions through a systematic review.

Methods: PubMed, PsycINFO, Scopus, UK Trial Database, ClinicalTrials.gov, and the Cochrane Library were surveyed up to May 2018. All types of clinical studies were considered. A total of 2 independent reviewers assessed quality using Scottish Intercollegiate Guidelines Network (SIGN) criteria. Ratings were used to provide an overall score for each study (low, moderate, or high). Data were synthesized in evidence tables.

Results: From 233 potentially relevant publications, only 28 studies were included. Of these, 13 (46%) were randomized control trials, 11 were single-arm studies (39%), 3 were nonrandomized controlled trials (11%), and 1 study was a cluster randomized trial (4%). The studies were classified as low (15), high (7), and moderate (6) quality according to SIGN criteria. All studies focused on efficacy, with only 1 trial mentioning safety and another 1 effectiveness. In 11 studies, the apps were used as stand-alone interventions, the others were multicomponent studies that included other tools for support such as sensors or websites. The main

management tool included in the apps was feedback messaging (24), followed by goal-setting mechanisms (20) and self-monitoring (19). The majority of studies took weight or body mass index loss as the main outcome (22) followed by changes in physical activity (14) and diet (12). Regarding outputs, usability, adherence, and engagement (17) were the most reported, followed by satisfaction (7) and acceptability (4).

Conclusions: There is a remarkable heterogeneity among these studies and the majority have methodological limitations that leave considerable room for improvement. Further research is required to identify all relevant criteria for assessing the efficacy of mHealth interventions in the management of overweight and obesity.

Trial Registration: PROSPERO CRD42017056761; https://tinyurl.com/y2zhxtjx

(JMIR Mhealth Uhealth 2019;7(10):e12612) doi:10.2196/12612

KEYWORDS

mHealth; obesity; overweight; systematic review; technology assessment

Introduction

Background

Obesity and overweight are considered major public health concerns because of their high prevalence and association with various health complications including cardiovascular disease, type 2 diabetes, and cancer [1,2]. As the aspects that influence overweight and obesity are diverse—comprising individual, genetic, and environmental factors—their prevention and treatment are also complex. For a successful treatment, multifactorial approaches are required, with diet and exercise plans reinforced with psychological therapy and behavioral change strategies [3].

In recent years, we have witnessed a revolution in the use of apps within personal health care, as they are fast, flexible, handy, versatile, manageable, and illustrative tools that can empower patients. Hence, mobile health (mHealth) can play an important adjuvant role in the prevention and treatment of overweight and obesity by tracking physical activity (PA), enabling self-reporting of dietary patterns, providing recommendations to achieve healthier habits, guidance, advice, tips, and motivational strategies to achieve personalized goals; all are relevant aspects for the prevention and treatment of obesity, as recognized in numerous guidelines [3,4].

The Global Observatory for eHealth of the World Health Organization (WHO) defines mHealth as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [5]. The management—and in some cases the prevention—of chronic diseases has been one focus in recent developments in both electronic health and mobile health (eHealth and mHealth) [6]. There are over 325,000 health apps on the market, with the most downloaded being those relating to PA and weight control: that is, those that support a healthy lifestyle [7]. However, information on how the effectiveness, efficacy, and safety of mHealth apps in overweight and obesity management are assessed is severely lacking. It is important to note that according to mHealth publishers, over 53% of their health apps portfolio available in 2015 were downloaded less than 5000 times [7]. Evidence of the efficacy of mHealth app strategies in improving chronic health and well-being is mixed; therefore, while some mHealth interventions show promise in supporting weight management [8,9], others do not [10,11]. Numerous efforts to address this challenging issue are underway and some aspects that may be linked to a lack of efficacy have been identified. These are, among others, the poor quality of many apps, a lack of guidance on the usefulness of an app, and a low level of support from health professionals [12,13]. Should mHealth apps be rigorously evaluated to ensure they provide evidence-based effectiveness, safety, and efficacy? Up to now, mHealth evaluation methodology has not deviated from customary methods (mainly randomized controlled trials [RCTs]), despite claims that alternative, shorter, and more inexpensive design methods are required [14].

There are several initiatives attempting to define how apps should be evaluated. However, all of these consider only partial aspects of evaluation [15]. Although medical regulatory bodies have not validated the safety and quality of these technologies, individuals have adopted mHealth devices as self-management aids. However, medical professionals are often at a loss as to how to relate to them [16]. Owing to this rapid consumer-based introduction to the world of patient health aids, mHealth solutions present unique and stakeholder-specific challenges to the medical environment. Patients, health care providers, administrators, authorities, and mHealth developers alike are operating without a clear direction, which may lead to problems, including the improper use of mHealth interventions by individuals and the inability of medical systems to react due to a lack of technological and organizational support. Users and health care professionals should be aware of the quality of health apps they use or prescribe. The use of classic methodologies such as RCTs may not be the optimal procedure for evaluating all the dimensions of mHealth. Ideally, clinicians, health administrations, and users need instruments that enable the evaluation of e-interventions as a whole. From a global perspective, these instruments should facilitate the process of verification, validation, impact assessment, and certification that ought to be a requirement for all mHealth implementation.

This lack of rigorous evaluation is an increasing concern for health authorities. A number of recommendations to ensure a minimum quality of mHealth interventions have already been defined by the WHO Technical Evidence Review Group [17]. In addition, both the *Food and Drug Administration* [18] and the *European Commission* [19] have made several attempts to establish mHealth assessment and, where appropriate,

certification criteria. However, in such a continuously evolving field, it has been difficult to reach a consensus.

Objectives

The aim of this paper was to undertake a systematic review of efficacy, safety, and effectiveness assessment criteria in use, including both outputs and outcomes, to assess weight control, overweight, and obesity management in mHealth interventions. These criteria will later be included in a tool for assessing mHealth interventions intended to manage overweight and obesity.

Methods

This systematic review was prospectively registered with PROSPERO (CRD42017056761) [20]. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement was used as a guide for reporting this review [21]. Owing to the methodological and statistical heterogeneity of the included studies, a descriptive approach was adopted in the research synthesis.

Eligibility Criteria

Any trial that assessed the efficacy and/or safety and/or effectiveness of mHealth-based interventions for overweight or obesity management was considered. No restrictions in terms of target population were foreseen. We define efficacy as changes in lifestyles on the basis of diet and PA in a controlled population; effectiveness in the general population; and safety as a lack of adverse effects resulting from mHealth interventions. Studies carried out with less than 10 individuals were excluded. We assessed the quality of trials according to the Scottish Intercollegiate Guidelines Network (SIGN) criteria [22]. Taking the objective of this review into consideration, all studies were included regardless of quality.

Information Sources

A systematic search was conducted in the following databases: MEDLINE, EMBASE, PsycINFO, The Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials [CENTRAL]), UK Trial Database, and Scopus. This survey was supplemented through the snowballing technique to identify relevant articles in the references of those returned by the search. A manual search was also conducted on the indices of the following publications: *Journal of Medical Internet Research* and *JMIR mHealth and uHealth*. The survey period included all articles published up to May 2018. All types of clinical studies published in English, French, or Spanish were considered.

Search Strategy

The search strategy included both controlled vocabulary and free-text terms. The terms used were apps, mHealth, eHealth, overweight, obesity, efficacy, security, safety, effectiveness, and evaluation (see Multimedia Appendix 1).

Study Selection and Data Collection Process

All identified references were imported into Mendeley v1.18 (Elsevier) and duplicates eliminated. A total of 6 researchers undertook the review process, which was conducted in 2 stages. First, each article identified was randomly assigned to 2 reviewers to independently review the title and abstract. Articles that met the inclusion criteria were full-text reviewed and quality-assessed by 2 independent reviewers. In cases of disagreement, a third reviewer made the final decision. Study features and outcomes were entered into a database specifically designed for this review. Risk of bias was assessed according to SIGN codes for study assessment [22]. Those trials that were clearly of an adequate quality were graded as *high or* ++ (very low risk of bias) or *moderate or* + (low risk of bias), while those of insufficient quality were graded as *low or*–(high risk of bias).

Results

Selection of Studies

A total of 233 potentially relevant publications (17 from a manual search) were identified as eligible. From these, 19.7% (46/233) were identified as duplicates. From the remaining (187), only 49.2% (92/187) were accepted for abstract review. Out of these, 47% (44/92) were excluded for not following inclusion criteria. A full-text review was conducted on 48 studies. After peer review, 30 articles corresponding to 28 different studies (62.5% from the total included for full-text review) were finally included in this nonquantitative review. The exclusion criteria were as follows: published study protocols (n=8), out of scope studies that were not using an mHealth intervention (n=4), or those studies in which final outcomes were other than efficacy or safety (n=6; Figure 1; see Multimedia Appendix 2).

The main characteristics of the 28 studies included are detailed in Multimedia Appendix 3. Studies appear in alphabetical order of the first author within chronological years. All selected studies focus on efficacy; only 1 of them assesses effectiveness and 1 also focused on safety, although this was not the main outcome of the study.

In total, 46% of the studies (13/28) are RCTs, 1 is a cluster-randomized trial [23], 3 are nonrandomized controlled trials, and the remainder (11) single-arm trials; 2 of the RCTs include more than 1 intervention. Carter et al [24] studied the efficacy of a smartphone app or website in self-monitored weight management, and Hurkmans et al [25] compared 1 stand-alone app intervention with face-to-face and blended interventions. All studies compare pre and post outcomes to analyze the intervention's efficacy. According to SIGN criteria, the majority of studies are of low (15) or moderate (6) quality, with only 7 studies reaching high quality. A low quality rating most often resulted from small sample size, inadequate length of study, or possible selection and information bias.

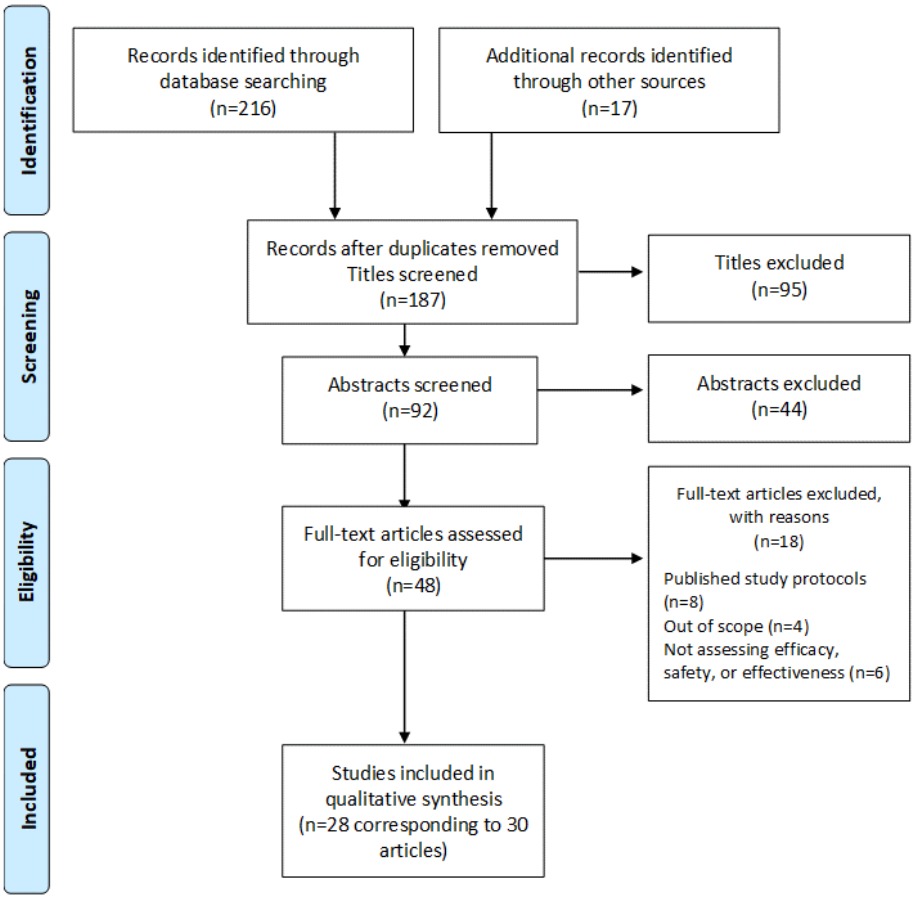
The number of participants ranged from 10 to 1012, but most studies (17) covered less than 100 people. One trial [26] had 15,310 participants, but the majority (83%) remained nonactive during the intervention. Most studies had a majority of adult women; in 6, all participants were women. There are also 4 studies targeted at children and teens. Most trials were targeted at people with overweight or obesity but no other health condition: exceptions targeted people with a severe mental illness [27], heart disease [28], type 2 diabetes [29] or

prediabetes [30,28], cancer survivors [31,32], and people with metabolic syndrome [33].

Apart from one 24-month trial [34], the studies were conducted over short periods of time, ranging from 3 weeks to 6 months.

The countries where the studies were carried out were the United States (17), Australia (3), Korea (2), the United Kingdom, Belgium, Spain, The Netherlands, China, and Israel.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram of selection of papers for inclusion in the review.



Elements Included in the Mobile Health Interventions

In regard to the specificities of mHealth interventions (Multimedia Appendix 4), only 39% (10 out of 28) focused on a specific stand-alone app, with the majority addressing multicomponent interventions—including armband sensors, pedometers, wireless scales, and other monitoring devices, or websites—for weight management, intended to increase PA, reduce sedentary habits, and/or improve dietary patterns. The most common elements included in the trials were the receiving of feedback messages (24 studies out of 28, 85%), setting of goals (20), and self-monitoring (19). These feedback messages could be personalized reminders, recommendations based on the self-monitoring, standard counseling or health coach counseling through the app, and/or a more synchronic intervention. New elements have been introduced to mHealth interventions in recent years, such as gamification [35,30,23,29,26], entertainment aspects [30,36,37,23,38], and peer contact through community blogs [39] or virtual teams [30] on social networks [37,27,40] such as Facebook [27,40,25] and We Chat [26]. It is worth mentioning that only specific frameworks were mentioned when defining strategies for behavioral change, such as the transtheoretical model of behavior change for TXT2BFiT [39] and CITY [34], intervention and Control Systems Theory for eBalance App [41], self-regulation theory for Balance It intervention [23], and increasing adherence, such as Mechanics-Dynamics-Aesthetics for With U App [40]. Social Cognitive Theory is the one most often referred to by the Vegethon app [42], Loose It app [31], and Alive-PD [30] for which several other frameworks were also considered: behavioral economics, positive psychology, and the theory of planned behavior. One study was based on the Diabetes Prevention Program [43] and LookAHEAD (Action for Health in Diabetes) trials [44]. One study was based on an addiction treatment approach [37].

Output Tools and Measures

Although their main aim was to measure the efficacy of mHealth interventions, most of the selected studies also measured other outputs that might be relevant to determine primary outcome measures (23 out of 28, 82%). Multimedia Appendix 5 [45,46-50] shows the outputs and the main tools used to measure them.

Acceptability

A total of 4 studies out of 28 (14%) attempted to measure participants' acceptance of the intervention, using mixed methods (survey, focus groups, and data performance tracking) [24,36,38,32]. Results showed that participants are willing to participate in these interventions, although receiving a smartphone [24] or doing it on a voluntary basis are elements that should be considered [32].

Usability/Adherence/Engagement

These 3 dimensions have been considered together, as the main analysis strategies used (data tracking and surveys) integrate all 3 aspects. Only 1 study used a validated questionnaire to assess usability [40], the System Usability Scale questionnaire. Several studies measured these outputs through different strategies, mainly data tracking. Results were very heterogeneous and depend on the study design and the specificities of each intervention.

Satisfaction

Only 7 studies analyzed the satisfaction rate of users [51,33,37,41,52,40,53], with 3 of these using standardized validated tools [54-56]. Results showed that most of the participants were very satisfied with the intervention, although a few considered the app too tedious to use.

Motivation to Lose Weight and to Continue the Intervention

Few studies [57,40] addressed continued motivation to lose weight after the intervention [41]. Only 1 of these [40] used a previously validated methodology, whereas the other 2 studies assessed motivation or intention to continue through a Likert scale [57] and self-reported questionnaires [41]. Results showed increases in users' motivations and in the adoption of a positive attitude toward managing their overweight or obesity.

Perceived Peer Support

Out of the 7 studies dealing with peer support, only 2 attempted to assess the perception of this support [27,32]. Both showed a high perceived importance of peer support in reducing stress associated with the intervention.

Outcome Tools and Measures

The end point outcomes of the selected studies were as follows: reduction of weight and body mass index (BMI) as well as fat mass and waist and hip circumferences; changes in dietary habits, PA, and screen time patterns; biochemical measurements; and blood pressure (Multimedia Appendix 6 [58-73]).

Weight and Body Mass Index

Most of the studies (22/28, 78%) considered reduction of weight and/or BMI as the main outcome with which to assess intervention efficacy. Devices used to measure weight and/or height were detailed in 17 trials, and only a few relied on self-reported data [74,26,53]. Partridge et al [74] did not report any differences between self-reported data and scale measures. All trials measured reduction in body weight, but in 3 studies [34,75,26] there were no differences between control and intervention groups; 3 other studies [35,41,25] noted differences between groups, but if statistical significance is taken as $P<.05$, these did not reach the threshold. Interventions that included face-to-face elements produced significantly better final outcomes [34,25]; 5 other 2-arm trials showed a clear and statistically significant reduction in body weight [24,76,74,53]. All pretest-posttest single-arm studies also measured weight reduction after intervention, but this was not always significant [38,52,32]; in one of these studies, considered to be of low quality, all weight was fully regained by 24 weeks after the intervention [52].

Fat Mass

Fat mass reduction was measured in 3 studies [35,33,40] through bioelectrical impedance and producing controversial results. In the 2 RCTs [35,33], fat reduction was statistically significant when comparing the control and intervention groups. In a single-arm trial [40], reduction was not statistically significant.

Waist and Hip Circumferences

Fukuoka et al [76] measured changes in hip circumference, noting significant changes in the intervention group. In total, 8 trials [30,31,41,38,40,77,26,25] measured changes in waist circumference, although the protocols in use varied or were not clearly specified. Results were controversial. Safran et al [41] and He et al [26] reported no changes, whereas 5 trials [30,31,38,40,77] identified a clear and significant reduction in waist circumference, whereas Hurkmans et al [25] recorded nonsignificant reductions.

Dietary Pattern

We identified 12 trials that assessed changes in dietary patterns [78,76,42,25,31-39,41-34,32,79]. All trials employed 2-arm pretest-posttest analysis, except for Quinitliani et al [32] and McCarroll et al [31]. Only 3 [31,33,23] did not use validated and previously published tests or questionnaires. A total of 6 studies [31-39,23,34,79] found no change when comparing fruit and vegetable consumption or the macronutrient composition of daily diet between 2 groups, although the intervention group appeared to adhere more closely to a Mediterranean diet [79] or were more likely to consume vegetables [39]. Other studies were able to demonstrate a clear improvement in dietary patterns: Fukuoka et al [76] observed a clear decrease in the intake of saturated fat; Mumah et al [42] identified a higher intake of vegetables; Safran et al [41] observed an improvement in diet quality; and Hurkmans et al [25] noted a clear and significant decrease in total energy intake. Both Nollen et al [78] and Quintiliani et al [32] perceived a statistically insignificant increase in fruit and vegetable consumption. In regard to sugar-sweetened beverages, 2 studies were able to measure a significant [76,39] or slight decrease [78]. Unexpectedly, participants in the single-arm study by Quintiliani et al [32] consumed more sugar-sweetened beverages after the intervention.

Physical Activity Pattern

In total, 14 of the 28 studies (50%) had PA pattern as a main end point. Various strategies were used to measure PA: (1) data tracking through accelerometers [36,79,25], pedometers [76], armband sensors [57,36], or logs from the apps [31]; (2) standard questionnaires such as International Physical Activity

Questionnaire (IPAQ) or IPAQ-Short Form (IPAQ-SF) [33-74,32], the Paffenbarger Physical Activity Questionnaire [34], or a modification of the IPAQ questionnaire [41]; (3) semistructured interviews [79]; and (4) ad hoc questionnaires [23]. The most common measurements were daily number of steps [36-28,79,25] and time spent doing PA [57,36,31,74,41,27]. The number of metabolic equivalents of task [33,32] and weekly self-reported spent kilocalories [34] were also used.

All studies except 4 [33,23,34,79] showed an improvement in PA patterns. However, only 5 of these stated that the improvement was statistically significant [57,76,31,74,41].

Emotional Well-Being

As the intervention assessed was based on an addiction treatment approach, Pretlow et al [37] analyzed changes in self-esteem and the likelihood of turning to food when feeling stressed. They reported a significant improvement in self-esteem and control of participants' eating. The McCarroll trial [31] analyzed changes in quality of life for cancer survivors. There were no differences before and after the mHealth intervention.

Screen Time

Nollen et al [78] studied possible changes in screen time but recorded no differences between the control and intervention groups.

Biochemical Measurements

As blood fasting lipids and glucose levels are usually high among people with overweight and obesity; 5 studies included these as secondary outcomes. Only Block et al [30] could report a significant improvement in triglyceride/high-density lipoprotein ratio; 2 studies [79,25] showed a trend toward reduction but the results were not significant. The other 2 trials did not measure any change in either fasting lipids or glucose [76,33].

Blood Pressure

Fukuoka [76], Willey [77], and Mao [53] tracked changes in blood pressure as a secondary outcome. The 3 trials were able to measure significant reductions in both systolic blood pressure and diastolic blood pressure.

Safety

One high-quality trial [33] considered safety as an outcome to be measured. The aim of this study was to evaluate the effect of SmartCare intervention in patients with metabolic syndrome. They identified a number of mildly adverse events (14.2% in the intervention group and 13.3% in the control group). There were also serious adverse events: 1.4% corresponding to 3 cases in the intervention group, including 1 ankle fracture; and 2.4% (5 cases) in the control group, including dislocated vertebra, stress urinary incontinence, and the need for a knee operation.

Effectiveness

Only 1 study was targeted at the general population. He et al [26] conducted a low-quality trial on 15,310 people. No differences between the intervention and control group were shown in terms of losing weight.

Discussion

Principal Findings

In this systematic review, we have identified the range of dimensions and tools used to assess the efficacy of mHealth interventions intended to manage overweight and obesity. We have provided a descriptive analysis of 28 clinical trials along with an account of the components and elements included in each intervention. Outputs and outcomes used for the evaluation of trials have also been identified. No specific criteria for assessing safety or effectiveness have been identified due to the small number of studies focused on these aspects.

Our results show that researchers use the following primary end points to measure a study's success: (1) reduction in weight and/or BMI; (2) reduction in fat mass; (3) reduction in waist and hip circumference; (4) improvement in dietary habits/patterns; (5) increase in PA; (6) increase in emotional well-being; (7) decrease in screen time patterns; (8) improvement in biochemical measures; and (9) decrease in blood pressure. All these factors are closely linked to obesity and overweight and are risk factors for future chronic disease. Although the main aim of most of the studies was to measure the efficacy of mHealth interventions, they also measured other outputs that might be relevant for determining the success of the intervention, such as (1) acceptability, (2) adherence, usability, and engagement, (3) satisfaction, (4) motivation, (5) intention to continue, and (6) perceived support. All these aspects appear to affect whether an intervention will be successful. Tests and questionnaires are the most prevalent tools used for assessment, whether existing and previously validated or devised for the situation. Objective data tracking of PA performance through the mHealth-based intervention, when possible, was a common strategy for avoiding self-reported data. It appears to be highly important to gather objective data and use standardized protocols when assessing the usability and efficacy of mHealth interventions. The mHealth strategies considered to be more sophisticated usually include a higher number of elements. Although the recent strategies of peer support and gamification appear to improve efficacy by increasing engagement and motivation, there is as yet not enough evidence to state this definitively.

The acknowledgement and evaluation of comprehensive sociodemographic differences, such as race/ethnicity, socioeconomic status, and sex, are severely lacking. Future analyses of mHealth interventions should consider and, whenever possible, include eHealth literacy aspects in an effort to reduce communication inequalities across groups [80]. Unless designers and developers of health care information technologies address security challenges, benefits from health care information technology will be scarce [81]. Another aspect we have found to be lacking from mHealth evaluation studies is assessment of clinical data confidentiality.

Previously published reviews have concluded that despite a lack of evidence concerning the best use of technology in weight loss interventions, when the optimal combination of technological components is determined, technology-based interventions will be a valid tool for weight loss [82]. Others

have been less optimistic and feel that future studies must use larger study samples, longer interventions, and follow-up periods [83]. One meta-systematic review concluded that despite the increasing popularity of mHealth, evidence for efficacy is still limited due to the low methodological quality of research [84]. We believe the issue may be how mHealth strategies are assessed and validated: this cannot be carried out in the same manner as research into drugs, and more adapted and/or flexible approaches are needed to explore new evaluation tools. An instrument intended to evaluate mHealth should include verification of its scientific content and mechanisms that ensure data privacy as well as safe usage. Verification of these aspects would ideally be mandatory before release and use in clinical practice. In the second phase, evaluations of effectiveness, efficacy, and usability should include user feedback, and adaptability and cost-effectiveness should also be addressed. This second phase of evaluation could be quantitative, enabling assessment of an mHealth intervention's quality and comparison with others.

Currently, most apps used or prescribed in daily clinical practices have only received technical verification or partial clinical validation on the basis of a small group of patients.

Future research is necessary to better assess mHealth interventions in development and before clinical application. It is important to find a balance between the necessary development of mHealth, which should be characterized as disruptive, innovative, and rapid, and the imperative need to validate mHealth interventions. From a Public Health point of view, it is necessary to avoid or minimize the potential problems a new mHealth intervention might create without accurate evaluation. It has been argued that the app market regulates itself: the good persist; the bad disappear. However, in such a potentially harmful field as mHealth, there is a need for new approaches and tools, and a multidisciplinary assessment process [14,85-89].

Limitations

One of the main limitations of this review is publication bias. References from other sources such as conferences and meetings have not been included. Although the number of scientific journals that publish mHealth-related articles has increased in recent years, there is a lot of gray literature surrounding this field that we may have missed. Moreover, only studies published in English, French, or Spanish have been included. The heterogeneity of interventions and populations and the low number of participants in many studies have made it difficult to synthesize results. Most of the studies included were deemed to be of moderate-low quality, and consequently findings need to be considered with caution. In total, 11 studies lacked a control group and therefore results cannot be attributable to the technology-based intervention alone. One must also take into account the established fact that individuals who agree to participate in intervention studies have greater motivation to change their lifestyles than the general population.

Finally, only 1 study was identified with the primary aim of assessing the safety and effectiveness of an mHealth intervention. Given awareness of safety-related issues such as a possible increase in anxiety and stress due to the use of mHealth intervention and the possible promotion of eating disorders, this is rather surprising. Furthermore, the studies reviewed largely assessed dietary habits and PA, ignoring other possible outcomes relating to body weight such as sleeping behavior. This also needs to be addressed in future research.

Conclusions

The potential for apps to positively help users manage their obesity or overweight has yet to be attained. Studies assessing the success of mHealth interventions are remarkably heterogeneous and most have methodological limitations that leave significant room for improvement regarding quality. Further research is needed to identify all relevant criteria for assessing the efficacy of mHealth interventions in the prevention and management of overweight and obesity.

Acknowledgments

All authors contributed equally. The research for this paper was fully funded by the Instituto de Salud Carlos III from the Spanish Ministry of Science, Innovation and Universities, grant number PI16/01764 co-funded by FEDER.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 35 KB - mhealth_v7i10e12612_app1.pdf]

Multimedia Appendix 2
Excluded publications.
[PDF File (Adobe PDF File), 181 KB - mhealth_v7i10e12612_app2.pdf]

Multimedia Appendix 3
Characteristics of the selected studies.

[PDF File (Adobe PDF File), 253 KB - mhealth_v7i10e12612_app3.pdf]

Multimedia Appendix 4
Elements included in the mobile health interventions of the selected studies.
[PDF File (Adobe PDF File), 229 KB - mhealth_v7i10e12612_app4.pdf]

Multimedia Appendix 5
Output tools and results from the selected studies.
[PDF File (Adobe PDF File), 242 KB - mhealth_v7i10e12612_app5.pdf]

Multimedia Appendix 6
Main outcome results from the selected studies.
[PDF File (Adobe PDF File), 291 KB - mhealth_v7i10e12612_app6.pdf]

References

1. Garg SK, Maurer H, Reed K, Selagamsetty R. Diabetes and cancer: two diseases with obesity as a common risk factor. Diabetes Obes Metab 2014 Feb;16(2):97-110 [FREE Full text] [doi: 10.1111/dom.12124] [Medline: 23668396]
2. Gittelsohn J, Trude A. Diabetes and obesity prevention: changing the food environment in low-income settings. Nutr Rev 2017 Jan;75(suppl 1):62-69 [FREE Full text] [doi: 10.1093/nutrit/nuw038] [Medline: 28049750]
3. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
4. Aguilar-Martínez A, Solé-Sedeño JM, Mancebo-Moreno G, Medina FX, Carreras-Collado R, Saigí-Rubió F. Use of mobile phones as a tool for weight loss: a systematic review. J Telemed Telecare 2014 Sep;20(6):339-349. [doi: 10.1177/1357633X14537777] [Medline: 24875928]
5. World Health Organization. Mhealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on Ehealth. Geneva, Switzerland: World Health Organization; 2011.
6. McKinstry B, Hanley J, Wild S, Pagliari C, Paterson M, Lewis S, et al. Telemonitoring based service redesign for the management of uncontrolled hypertension: multicentre randomised controlled trial. Br Med J 2013 May 24;346:f3030 [FREE Full text] [doi: 10.1136/bmj.f3030] [Medline: 23709583]
7. Research2Guidance. 2016. mHealth Economics 2016 – Current Status and Trends of the mHealth App Market URL: https://research2guidance.com/product/mhealth-app-developer-economics-2016/ [accessed 2017-02-17]
8. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
9. Ganesan AN, Louise J, Horsfall M, Bilsborough SA, Hendriks J, McGavigan AD, et al. International mobile-health intervention on physical activity, sitting, and weight: the Stepathlon cardiovascular health study. J Am Coll Cardiol 2016 May 31;67(21):2453-2463 [FREE Full text] [doi: 10.1016/j.jacc.2016.03.472] [Medline: 27050185]
10. Laing BY, Mangione CM, Tseng CH, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
11. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
12. Main C, Moxham T, Wyatt JC, Kay J, Anderson R, Stein K. Computerised decision support systems in order communication for diagnostic, screening or monitoring test ordering: systematic reviews of the effects and cost-effectiveness of systems. Health Technol Assess 2010 Oct;14(48):1-227 [FREE Full text] [doi: 10.3310/hta14480] [Medline: 21034668]
13. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
14. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mhealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
15. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth assessment: conceptualization of a global framework. JMIR Mhealth Uhealth 2017 May 2;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
16. World Health Organization. Tuberculosis (TB): Frequently Asked Questions on Global Task Force on Digital Health for TB and Its Work URL: http://www.who.int/tb/areas-of-work/digital-health/faq/en/ [accessed 2017-02-27] [WebCite Cache ID 73MSOg9b3]

17. 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]
18. Food and Drug Administration. 2015. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/mobile-medical-applications [accessed 2017-02-27]
19. European Commission. 2014. Green Paper on Mobile Health ('mHealth') URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2017-02-27]
20. Carrion C, Garcia-Lorda P, Zamora A, Paluzié G, Moharra M, Puigdomènech E. PROSPERO - University of York. 2018. Systematic Review and Meta-analysis of Clinical Trials Assessing Efficacy, Effectiveness and Security of Overweight and Obesity Management Apps URL: http://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017056761 [accessed 2019-09-09]
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
22. Scottish Intercollegiate Guidelines Network (SIGN). 2015. SIGN 50: A Guideline Developer's Handbook URL: https://www.sign.ac.uk/assets/sign50_2015.pdf [accessed 2017-02-26]
23. Spook J, Paulussen T, Kok G, van Empelen P. Evaluation of a serious self-regulation game intervention for overweight-related behaviors ('Balance It'): a pilot study. J Med Internet Res 2016 Sep 26;18(9):e225 [FREE Full text] [doi: 10.2196/jmir.4964] [Medline: 27670222]
24. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
25. Hurkmans E, Matthys C, Bogaerts A, Scheys L, Devloo K, Seghers J. Face-to-face versus mobile versus blended weight loss program: randomized clinical trial. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e14 [FREE Full text] [doi: 10.2196/mhealth.7713] [Medline: 29326093]
26. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
27. Aschbrenner KA, Naslund JA, Shevenell M, Kinney E, Bartels SJ. A pilot study of a peer-group lifestyle intervention enhanced with mhealth technology and social media for adults with serious mental illness. J Nerv Ment Dis 2016 Jun;204(6):483-486 [FREE Full text] [doi: 10.1097/NMD.0000000000000530] [Medline: 27233056]
28. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 9;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
29. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile diabetes prevention program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
30. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
31. McCarroll ML, Armbruster S, Pohle-Krauza RJ, Lyzen AM, Min S, Nash DW, et al. Feasibility of a lifestyle intervention for overweight/obese endometrial and breast cancer survivors using an interactive mobile application. Gynecol Oncol 2015 Jun;137(3):508-515. [doi: 10.1016/j.ygyno.2014.12.025] [Medline: 25681782]
32. 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;2(1):e4 [FREE Full text] [doi: 10.2196/cancer.5305] [Medline: 27761518]
33. Oh B, Cho B, Han MK, Choi H, Lee MN, Kang HC, et al. The effectiveness of mobile phone-based care for weight control in metabolic syndrome patients: randomized controlled trial. JMIR Mhealth Uhealth 2015 Aug 20;3(3):e83 [FREE Full text] [doi: 10.2196/mhealth.4222] [Medline: 26293568]
34. Svetkey LP, Batch BC, Lin P, Intille SS, Corsino L, Tyson CC, et al. Cell phone intervention for you (CITY): a randomized, controlled trial of behavioral weight loss intervention for young adults using mobile technology. Obesity (Silver Spring) 2015 Nov;23(11):2133-2141 [FREE Full text] [doi: 10.1002/oby.21226] [Medline: 26530929]
35. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
36. Finkelstein J, Bedra M, Li X, Wood J, Ouyang P. Mobile app to reduce inactivity in sedentary overweight women. Stud Health Technol Inform 2015;216:89-92. [Medline: 26262016]

37. Pretlow RA, Stock CM, Allison S, Roeger L. Treatment of child/adolescent obesity using the addiction model: a smartphone app pilot study. Child Obes 2015 Jun;11(3):248-259 [FREE Full text] [doi: 10.1089/chi.2014.0124] [Medline: 25760813]
38. Hutchesson MJ, Morgan PJ, Callister R, Pranata I, Skinner G, Collins CE. Be Positive Be Healthe: development and implementation of a targeted e-health weight loss program for young women. Telemed J E Health 2016 Jun;22(6):519-528 [FREE Full text] [doi: 10.1089/tmj.2015.0085] [Medline: 26701611]
39. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mhealth lifestyle program with telephone support (TXT2BFiT) to prevent unhealthy weight gain in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2015 Jun 15;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
40. Lee J, Kim J. Development and efficacy testing of a social network-based competitive application for weight loss. Telemed J E Health 2016 May;22(5):410-418. [doi: 10.1089/tmj.2015.0067] [Medline: 26540485]
41. Naimark JS, Madar Z, Shahar DR. The impact of a web-based app (eBalance) in promoting healthy lifestyles: randomized controlled trial. J Med Internet Res 2015 Mar 2;17(3):e56 [FREE Full text] [doi: 10.2196/jmir.3682] [Medline: 25732936]
42. Mummah S, Robinson TN, Mathur M, Farzinkhou S, Sutton S, Gardner CD. Effect of a mobile app intervention on vegetable consumption in overweight adults: a randomized controlled trial. Int J Behav Nutr Phys Act 2017 Sep 15;14(1):125 [FREE Full text] [doi: 10.1186/s12966-017-0563-2] [Medline: 28915825]
43. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]
44. Look AHEAD Research Group, Wadden TA, West DS, Delahanty L, Jakicic J, Rejeski J, et al. The Look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. Obesity (Silver Spring) 2006 May;14(5):737-752 [FREE Full text] [doi: 10.1038/oby.2006.84] [Medline: 16855180]
45. Sauro J. MeasuringU. 2011. Measuring Usability With the System Usability Scale (SUS) URL: https://measuringu.com/sus/ [accessed 2015-09-20] [WebCite Cache ID 73NdoZEYg]
46. Jung Y. No verification on the participation behavior model of participants in leisure sport and exercise. Korean J Sport Psychol 2008;19:195-214.
47. Yu J. The Relationship Between Fun Factor, Exercise Immersion and Participation in Women's Leisure Dance. In: Graduate School of Kyunghee University. Seoul, South Korea: Graduate School of Kyunghee University; 2011.
48. Parmenter K, Wardle J. Development of a general nutrition knowledge questionnaire for adults. Eur J Clin Nutr 1999 Apr;53(4):298-308. [Medline: 10334656]
49. Caron J. [A validation of the social provisions scale: the SPS-10 items]. Sante Ment Que 2013;38(1):297-318 [FREE Full text] [doi: 10.7202/1019198ar] [Medline: 24337002]
50. World Health Organization. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:i-xii, 1-253. [Medline: 11234459]
51. Thomas JG, Wing RR. Health-e-call, a smartphone-assisted behavioral obesity treatment: pilot study. JMIR Mhealth Uhealth 2013 Apr 17;1(1):e3 [FREE Full text] [doi: 10.2196/mhealth.2164] [Medline: 25100672]
52. Jensen CD, Duncombe KM, Lott MA, Hunsaker SL, Duraccio KM, Woolford SJ. An evaluation of a smartphone-assisted behavioral weight control intervention for adolescents: pilot study. JMIR Mhealth Uhealth 2016 Aug 23;4(3):e102 [FREE Full text] [doi: 10.2196/mhealth.6034] [Medline: 27554704]
53. Mao AY, Chen C, Magana C, Barajas KC, Olayiwola JN. A mobile phone-based health coaching intervention for weight loss and blood pressure reduction in a national payer population: a retrospective study. JMIR Mhealth Uhealth 2017 Jun 8;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]
54. Shahar DR, Henkin Y, Rozen GS, Adler D, Levy O, Safra C, et al. A controlled intervention study of changing health-providers' attitudes toward personal lifestyle habits and health-promotion skills. Nutrition 2009 May;25(5):532-539. [doi: 10.1016/j.nut.2008.11.020] [Medline: 19230614]
55. Attkisson CC, Greenfield TK. The UCSF client satisfaction scales: I. The client satisfaction questionnaire-8. In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2004:799-811.
56. Lim SA, Kang SE. Development and validation study of the achievement motivation scale. Korean J Educ Psychol 2013;27:575-593.
57. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLoS One 2014;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]
58. 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]
59. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011 Oct 21;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]

60. Paffenbarger RS, Hyde RT, Wing AL, Hsieh CC. Physical activity, all-cause mortality, and longevity of college alumni. N Engl J Med 1986 Mar 6;314(10):605-613. [doi: 10.1056/NEJM198603063141003] [Medline: 3945246]
61. Larsson UE, Reynisdottir S. The six-minute walk test in outpatients with obesity: reproducibility and known group validity. Physiother Res Int 2008 Jun;13(2):84-93. [doi: 10.1002/pri.398] [Medline: 18446882]
62. Smith KJ, McNaughton SA, Gall SL, Blizzard L, Dwyer T, Venn AJ. Takeaway food consumption and its associations with diet quality and abdominal obesity: a cross-sectional study of young adults. Int J Behav Nutr Phys Act 2009 May 28;6:29 [FREE Full text] [doi: 10.1186/1479-5868-6-29] [Medline: 19473547]
63. Resnicow K, Odom E, Wang T, Dudley WN, Mitchell D, Vaughan R, et al. Validation of three food frequency questionnaires and 24-hour recalls with serum carotenoid levels in a sample of African-American adults. Am J Epidemiol 2000 Dec 1;152(11):1072-1080. [doi: 10.1093/aje/152.11.1072] [Medline: 11117617]
64. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. J Clin Epidemiol 1990;43(12):1327-1335. [doi: 10.1016/0895-4356(90)90099-b] [Medline: 2254769]
65. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. Arch Intern Med 2008 Apr 14;168(7):713-720. [doi: 10.1001/archinte.168.7.713] [Medline: 18413553]
66. Australian Institute of Health and Welfare. 2001. Risk Factor Monitoring: Physical Activity, Diet and Body Weight: Results From the 2001 National Health Survey URL: https://www.aihw.gov.au/getmedia/5f919628-dd75-4f3e-95d1-1ef3e2e86edb/padbwdb01.pdf.aspx?inline=true [accessed 2017-03-27]
67. Rifas-Shiman SL, Willett WC, Lobb R, Kotch J, Dart C, Gillman MW. PrimeScreen, a brief dietary screening tool: reproducibility and comparability with both a longer food frequency questionnaire and biomarkers. Public Health Nutr 2001 Apr;4(2):249-254. [doi: 10.1079/PHN200061] [Medline: 11299098]
68. Hedrick VE, Savla J, Comber DL, Flack KD, Estabrooks PA, Nsiah-Kumi PA, et al. Development of a brief questionnaire to assess habitual beverage intake (BEVQ-15): sugar-sweetened beverages and total beverage energy intake. J Acad Nutr Diet 2012 Jun;112(6):840-849 [FREE Full text] [doi: 10.1016/j.jand.2012.01.023] [Medline: 22709811]
69. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011 Jun;141(6):1140-1145. [doi: 10.3945/jn.110.135566] [Medline: 21508208]
70. Nurses' Health Study. 2003. Nurses Health Study II Questionnaire URL: https://www.nurseshealthstudy.org/sites/default/files/questionnaires/2001long.PDF [accessed 2017-02-18] [WebCite Cache ID 73Nktc61m]
71. Matthys C, Meulemans A, van der Schueren B. Development and validation of general FFQ for use in clinical practice. Ann Nutr Metab 2015;67:239 [FREE Full text]
72. 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]
73. Brucker PS, Yost K, Cashy J, Webster K, Cella D. General population and cancer patient norms for the functional assessment of cancer therapy-general (FACT-G). Eval Health Prof 2005 Jun;28(2):192-211. [doi: 10.1177/0163278705275341] [Medline: 15851773]
74. Partridge SR, Allman-Farinelli M, McGeechan K, Balestracci K, Wong AT, Hebden L, et al. Process evaluation of TXT2BFiT: a multi-component mHealth randomised controlled trial to prevent weight gain in young adults. Int J Behav Nutr Phys Act 2016 Jan 19;13:7 [FREE Full text] [doi: 10.1186/s12966-016-0329-2] [Medline: 26785637]
75. Gomez-Marcos MA, Patino-Alonso MC, Recio-Rodriguez JI, Agudo-Conde C, Romaguera-Bosch M, Magdalena-Gonzalez O, On Behalf the EVIDENT Investigators 11. Short- and long-term effectiveness of a smartphone application for improving measures of adiposity: a randomised clinical trial - EVIDENT II study. Eur J Cardiovasc Nurs 2018 Aug;17(6):552-562. [doi: 10.1177/1474515118761870] [Medline: 29488798]
76. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
77. Willey S, Walsh JK. Outcomes of a mobile health coaching platform: 12-week results of a single-arm longitudinal study. JMIR Mhealth Uhealth 2016 Jan 8;4(1):e3 [FREE Full text] [doi: 10.2196/mhealth.4933] [Medline: 26747611]
78. Nollen NL, Mayo MS, Carlson SE, Rapoff MA, Goggin KJ, Ellerbeck EF. Mobile technology for obesity prevention: a randomized pilot study in racial- and ethnic-minority girls. Am J Prev Med 2014 Apr;46(4):404-408 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.011] [Medline: 24650843]
79. Garcia-Ortiz L, Recio-Rodriguez JI, Agudo-Conde C, Patino-Alonso MC, Maderuelo-Fernandez J, Gento IR, EVIDENT Investigators Group, Mobilizing Minds Research Group. Long-term effectiveness of a smartphone app for improving healthy lifestyles in general population in primary care: randomized controlled trial (EVIDENT II study). JMIR Mhealth Uhealth 2018 Apr 27;6(4):e107 [FREE Full text] [doi: 10.2196/mhealth.9218] [Medline: 29702473]
80. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of ehealth usage: insights on the digital divide from the health information national trends survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]

81. Kotz D, Fu K, Gunter C, Rubin A. Security for mobile and cloud frontiers in healthcare. Commun ACM 2015 Jul 23;58(8):21-23. [doi: 10.1145/2790830]
82. Raaijmakers LC, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: a systematic review. Appetite 2015 Dec;95:138-151. [doi: 10.1016/j.appet.2015.07.008] [Medline: 26165415]
83. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mhealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
84. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
85. Dipankui MT, Gagnon MP, Desmartis M, Légaré F, Piron F, Gagnon J, et al. Evaluation of patient involvement in a health technology assessment. Int J Technol Assess Health Care 2015 Jan;31(3):166-170. [doi: 10.1017/S0266462315000240] [Medline: 26062904]
86. Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health 2014 Dec;104(12):e12-e22. [doi: 10.2105/AJPH.2014.302164] [Medline: 25322302]
87. Lilford RJ, Girling AJ, Sheikh A, Coleman JJ, Chilton PJ, Burn SL, et al. Protocol for evaluation of the cost-effectiveness of ePrescribing systems and candidate prototype for other related health information technologies. BMC Health Serv Res 2014 Jul 19;14:314 [FREE Full text] [doi: 10.1186/1472-6963-14-314] [Medline: 25038609]
88. Cortés MA, Cuenca MR, Verdugo RM, Cidoncha EC. High quantity but limited quality in healthcare applications intended for HIV-infected patients. Telemed J E Health 2014 Aug;20(8):729-735. [doi: 10.1089/tmj.2013.0262] [Medline: 24849001]
89. Nilsen W. American Association for the Advancement of Science. mHealth's Revolution: Balancing Help and Harm URL: https://www.aaas.org/sites/default/files/Nilsen%20mHealths%20Revolution%20Balancing%20Help%20and%20Harm.pdf [accessed 2017-03-15]

Abbreviations

BMI: body mass index
eHealth: electronic health
IPAQ: International Physical Activity Questionnaire
mHealth: mobile health
PA: physical activity
RCT: randomized control trial
SIGN: Scottish Intercollegiate Guidelines Network
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.10.18; peer-reviewed by J Alvarez Pitti, PH Lin, S Marsh; comments to author 04.04.19; revised version received 29.05.19; accepted 19.08.19; published 25.10.19.

Please cite as:
Puigdomenech Puig E, Robles N, Saigí-Rubió F, Zamora A, Moharra M, Paluzie G, Balfegó M, Cuatrecasas Cambra G, Garcia-Lorda P, Carrion C
Assessment of the Efficacy, Safety, and Effectiveness of Weight Control and Obesity Management Mobile Health Interventions: Systematic Review
JMIR Mhealth Uhealth 2019;7(10):e12612
URL: http://mhealth.jmir.org/2019/10/e12612/
doi:10.2196/12612
PMID:31654566

Original Paper

Engagement With a Digital Platform for Multimodal Cognitive Assessment and Multidomain Intervention in a Japanese Population: Pilot, Quasi-Experimental, Longitudinal Study

Jordan Glenn[1,2], MS, PhD; Erica Nicole Madero[1], MS, MPH; Michelle Gray[2], MS, PhD; Nami Fuseya[1], MSW; Mari Ikeda[3], BA; Tomoo Kawamura[3], MA; Yoshiko Arita[4], MA; Nick Thomas Bott[1,5], PhD

[1]Neurotrack Technologies Inc, Redwood City, CA, United States

[2]Exercise Science Research Center, Department of Health, Human Performance and Recreation, University of Arkansas, Fayetteville, AR, United States

[3]Nipponkoa Himawari Life Insurance, Inc, Tokyo, Japan

[4]Sompo Holdings, Tokyo, Japan

[5]Department of Medicine, School of Medicine, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Jordan Glenn, MS, PhD
Neurotrack Technologies Inc
399 Bradford St
Suite 101
Redwood City, CA, 94063
United States
Phone: 1 15615133865
Email: jordan@neurotrack.com

Abstract

Background: As the global prevalence of dementia continues to rise, multidomain lifestyle interventions that address modifiable risk factors associated with pathological cognitive decline are increasing. Although some digital options have been developed to increase the reach and scalability of these programs, because of cultural differences, the efficacy of the programs in one population cannot easily be generalized to populations in other countries.

Objective: This investigation aimed to examine the usability and engagement of a digitally delivered multidomain cognitive lifestyle intervention developed in the United States for a Japanese population.

Methods: This feasibility investigation utilized a quasi-experimental, single-arm, nonrandomized, longitudinal design where participants engaged in the behavioral intervention on a smartphone. Of the 559 participants that initially enrolled (age: mean 51 years, SD 7.5 years; 51.7% female [289/559]), 242 completed the final testing trial. Participants enrolled in a multidomain lifestyle program that consisted of (1) psychoeducational material, (2) physical activity tracker, (3) nutrition tracker, (4) audio-based meditations, and (5) health coaching. Engagement with the program was assessed through the total number of app sessions and the use of the exercise, diet, and meditation tracking features within the app. The total number of minutes exercised was collected through subjective user inputs, and nutrition was quantified by the Mediterranean-DASH Intervention for Neurodegenerative Delay diet adherence score.

Results: Significant relationships existed between overall nutrition score and frequency of nutrition tracking (r=0.18), frequency of physical activity tracking (r=0.19), and the total number of minutes exercised (r=0.22). Total minutes exercised was significantly correlated with total app sessions (r=0.57), frequency of physical activity tracking (r=0.85), frequency of nutrition tracking (r=0.64), number of times participants meditated (r=0.46), and total lessons read (r=0.36). The number of completed lessons was significantly related to frequency of physical activity tracking (r=0.40), frequency of nutrition tracking (r=0.43), the total number of times participants meditated (r=0.35), and total minutes exercised (r=0.33). Dividing the cohort into two groups based on lesson completion (<10 lessons completed vs ≥10 lessons completed), significant differences were observed between the total minutes exercised, frequency of physical activity tracking, frequency of nutrition tracking, and total number of times participants meditated (all P values <.01).

Conclusions: Overall, this cross-cultural feasibility study in Japanese users demonstrated that the various engagement metrics were significantly correlated, and greater engagement was related to improved nutrition scores and increased time exercising. In

addition, the relationships between lesson completion and other engagement metrics suggest that there may be value in exploring mechanisms that enhance lesson completion. Future research should examine the program in randomized control trials to more rigorously evaluate program efficacy.

(JMIR Mhealth Uhealth 2019;7(10):e15733) doi:10.2196/15733

KEYWORDS

cognitive decline; Alzheimer disease; lifestyle risk reduction; digital health; FINGER; neurocognitive tests; cognition; dementia

Introduction

Background

Characterized by declines in mental ability severe enough to interfere with daily life, Alzheimer disease (AD) poses a serious worldwide challenge as it relates to patients, their caregivers, and health care systems. Projections indicate that the global prevalence of AD is expected to triple to over 150 million individuals between 2015 and 2050 [1]. In the United States alone, AD costs are projected to grow from US $290 billion in 2019 to US $1.1 trillion in 2050, representing a 400% increase, whereas AD diagnoses are projected to increase by approximately 150%, from 5.5 million to 13.8 million over that same time span [2].

Although there is a massive cost associated with the disease once diagnosed, AD tends to go undetected for long periods because of a prolonged preclinical phase [3,4]. In this phase, neuronal and neurobiological changes can occur for years or decades before noticeable symptoms appear. Before the development of clinically detectable cognitive issues, people commonly experience a phenomenon termed as subjective cognitive decline (SCD) [5]. SCD is defined as subjective changes in memory and cognition that are perceived by the individual but are not associated with clinically measurable abnormalities [3,6]. Individuals experiencing SCD are considered at risk for developing dementia, specifically AD [7-9]. If left unchecked, SCD can evolve into mild cognitive impairment (MCI), an intermediate between normal cognitive function and diagnosable AD [6,7].

Age is the number one risk factor for AD, and the probability of being diagnosed with AD nearly doubles every 5 years after the age of 65 years [8]. Among adults older than 65 years, approximately 1 in 5 currently suffer from AD [9]. This number has increased dramatically in the United States since 2000, with an increase in diagnoses of 145% [9]. Although these numbers are alarming, they are dwarfed by the massive growth in Japan's aging population. Compared with the rest of the world, Japan has the largest aging population, with 33% of the population aged 60 years or older [10,11]. Furthermore, Japan's older population is projected to continue growing through 2050, reaching an unprecedented 42% of the population. This is critically important as the number of Japanese adults with dementia is estimated to be approximately 4.6 million, comprising nearly 15% of the older adult population [12]. When individuals with MCI are included, this number rises to approximately 8.6 million, constituting 30% of Japanese older adults [12]. The estimated cost of dementia in Japan in 2014, defined as the sum of costs for health care and formal and informal care, was approximately ¥14.5 trillion, or an estimated 3% of the nation's gross domestic product [13].

Through accumulated and ongoing randomized control trials, it is increasingly accepted that the development of AD is multifactorial, with a combination of modifiable and unmodifiable risk factors ultimately leading to the onset and progression of the disease [14-16]. As a result, it is difficult to conceptualize risk reduction for AD as a single-factor solution; rather, it is recommended to focus on multiple behavioral risk factors associated with disease development and progression [17]. The primary modifiable risk factors associated with AD are physical activity, diet, sleep, stress, social interaction, and cognitive engagement. Improvement in any of these factors may reduce the overall risk for disease onset and progression [16]. Furthermore, the ability to improve these risk factors at an individual level can have an overall beneficial effect on cognitive aging trajectories.

The nearly ubiquitous penetration of the wireless internet and subsequent adoption of smartphone technology offer an opportunity to address the demands and burdens associated with pathological cognitive decline by improving accessibility to digital solutions [18] and ultimately reducing associated health care costs [19]. To date, a number of multidomain interventions have focused on behavioral risk mitigation and enhancement of cognitive performance. These include the Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER) in Finland [16], US Study to Protect Brain Health Through Lifestyle Intervention to Reduce Risk in the United States [20], Singapore intervention study to prevent cognitive impairment and disability in Singapore [21], Maintain Your Brain in Australia [22], Multidomain Alzheimer's Prevention Trial (MAPT) in France [23], Prevention of Dementia by Intensive Vascular Care in the Netherlands [24], and all other studies under the scope of World Wide Fingers [25]. Although all these interventions contain Web-based components, none of them are completely digital, with the exception of Australia's Maintain Your Brain [22], thereby limiting the scalability and scope of the intervention's effectiveness. In fact, the only fully digital interventions focused on improving the aforementioned modifiable risk factors for pathological cognitive decline have been the Body Brain Life [26] and the Virtual Cognitive Health (VC Health) study [27,28], wherein users could access the entirety of the program via the internet through a laptop or desktop computer.

Objectives

On the basis of the results of the VC Health study, we developed the Neurotrack Memory Health Program (MHP) that transitions the form factor from a Web-based intervention to a smartphone.

Combining aspects of physical activity, diet, sleep, stress, social interaction, and cognitive engagement, this new method of program delivery expands the scope and accessibility of the behavioral intervention, which are limiting factors in receiving proper care [18]. Given the novelty of the program and method of delivery, our primary aim in this initial investigation was to evaluate the program's feasibility and acceptability of use in free-living individuals. On the basis of the current scope of the Japanese aging population and the disease-associated health care costs [12,13], this study investigated the feasibility and acceptability of the Neurotrack MHP in the Japanese population.

Methods

Study Overview

This feasibility investigation utilized a quasi-experimental, single-arm, nonrandomized, longitudinal (16-week) design in which participants used the Neurotrack MHP smartphone app to engage in a behavioral intervention designed to improve risk factors related to cognitive decline. As this program was originally an English product, there were requirements to make it accessible to Japanese populations, including translation to Japanese and adjustment of certain recommendations to be more culturally applicable (see below for a detailed description of the localization process). The smartphone app was used to collect all study data and monitor completion of program-related tasks. Participants completed all study procedures remotely via their smart device. The research protocol was approved for retrospective exemption by the institutional review board at the University of Arkansas for the analysis of deidentified data.

Participants

Participants in this investigation included employees and parents of employees of a large Japanese insurance company (Sompo Holdings, Tokyo, Japan); the program was offered as a free opportunity for all participants. It was important that this program was offered free of charge to remove incentivization bias on behalf of the participants. Participants were recruited through email outreach, flyers, and word of mouth throughout the company. As this was offered as an employee and relative benefit, the only requirements were that participants speak Japanese (to understand the program), be older than 40 years, and have access to a smartphone with internet access; no one was excluded from participation based on location or current health status. The rationale for including participants aged 40 years and older is based on epidemiological evidence suggesting that this is the age at which AD pathology begins to accumulate [29,30]. In addition, engaging in healthy habits in midlife may extend life spans by as many as 14 years in women and 12.2 years in men [31], and among middle-aged adults, 11.2% report SCD [32].

Localization for Japan

Once the English version of the program was finalized, an initial translation was provided by a bilingual (English and Japanese) speaker native to the Japanese language and culture. This was a requirement to ensure the nuances of the educational components were not lost in the translation process. After the program was translated, it was reviewed by another bilingual (English and Japanese) speaker native to the Japanese language and culture. After consensus was reached between each of the bilingual speakers regarding content, it was piloted with 5 individuals living in Japan who met the criteria for program delivery. Japanese individuals used for the pilot trial did not speak English to ensure a clear understanding of the content. After pilot data were collected, adjustments were made incorporating participant feedback. This process resulted in the creation of a fully Japanese version of the original English program.

Study Intervention

Once recruited, participants were sent an email to join the program, received an invitation to download the Neurotrack MHP app to their personal device, and were instructed to set up their in-app profile. To replicate the fully digital approach utilized in the VC Health study, participants were never required to report to a testing center or laboratory. After profile setup, participants completed a baseline survey to assess initial health status and behaviors (Table 1).

The survey questions were selected to understand the baseline risk factors commonly associated with AD (physical activity [33], diet [34], sleep [35], depression [36], and anxiety [37]).

Participants were then required to complete a visual paired comparison (VPC) eye-tracking assessment delivered through the Neurotrack MHP app [38,39]; a detailed description of the VPC task is included below. All other components of the program were blocked until after completion of the initial VPC test, which provided a baseline memory performance score (Figure 1).

After completion of the initial VPC assessment, participants had full access to the Neurotrack MHP. The Neurotrack MHP offered a number of primary functions that allowed users to track and engage in behaviors that may reduce the risk of cognitive decline. These features included (1) psychoeducational material, (2) self-reported physical activity tracker, (3) self-reported nutrition tracker, (4) audio-based meditations, and (5) health coaching. Each feature is described in detail below. These features were released on a timed schedule in an effort to gate the amount of information a participant was presented at any one time (Figure 2). Participants were encouraged to utilize the app daily to track their health behaviors and engage with the associated features. At the end of the 16-week intervention, all participants were provided access to a follow-up VPC assessment that could be taken again through the smartphone app.

Table 1. Baseline survey questionnaire provided to all participants. Note: the Patient Health Questionnaire-4 was used to assess depression and anxiety.

Question	Response type
On average, how many minutes of physical activity do you complete per week?	Numeric
During a normal week, how many days do you eat fish?	Numeric
During a normal week, how many days do you eat vegetables?	Numeric
During a normal week, how many days do you sleep more than 7 hours a night?	Numeric
Over the last 2 weeks, how often have you been bothered by the following problem: Feeling nervous anxious or on edge	Categorical (PHQ-4[a])
Over the last 2 weeks, how often have you been bothered by the following problem: Feeling down, depressed or hopeless	Categorical (PHQ-4)
Over the last 2 weeks, how often have you been bothered by the following problem: Not being able to stop or control worrying	Categorical (PHQ-4)
Over the last 2 weeks, how often have you been bothered by the following problem: Little interest or pleasure in doing things	Categorical (PHQ-4)

[a]PHQ-4: Patient Health Questionnaire-4.

Figure 1. Participant study funnel. MHP: Memory Health Program; VPC: visual paired comparison.

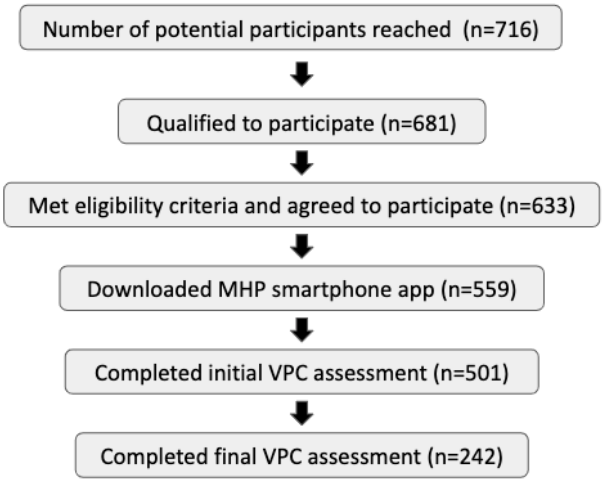
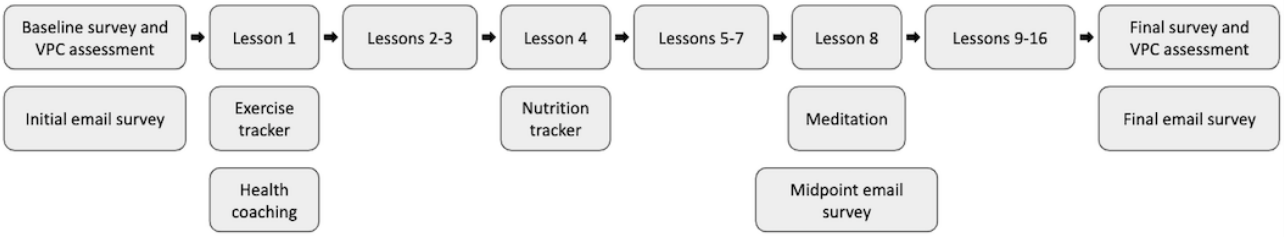


Figure 2. Timeline of the Memory Health Program and study protocol. VPC: visual paired comparison.



Intervention Components and App Functionality

Visual Paired Comparison Task

The VPC task was completed on the participant's device through the embedded Web camera. The construction of the VPC task has been explained in detail elsewhere [27,38,39]. Briefly, VPC tasks utilize eye-tracking technology to assess visual recognition memory by quantifying the time a participant spends viewing novel images as opposed to previously viewed images [39-41]. The VPC is administered in 2 separate phases: a familiarization phase and a testing phase.

During the assessment's familiarization phase, participants were presented with pairs of identical visual stimuli, each for 5 seconds. During the test phase, subjects were presented with pairs of visual stimuli (images) in the same format as the familiarization phase; however, this phase includes only 1 image from the familiarization phase and 1 novel image. The proportion of time a participant spent gazing at the novel image relative to the total viewing time produced a novelty preference score, with higher scores representing better declarative memory performance and lower scores indicating worse declarative memory performance [38,39]. Eye movements were tracked and scored. Detailed scoring information is published elsewhere [38].

Program Features and Functions

The psychoeducational material was organized into a series of 17 individual lessons (1 introductory lesson + 16 weekly informational lessons) that were presented to participants weekly. The content included information on the aforementioned lifestyle behaviors related to cognitive decline and general behavior change concepts such as goal setting and problem solving. Participants were asked to read the new lesson each week but were allowed to revisit the previous lessons at any point throughout the program. The physical activity tracker allowed participants to record their daily activities and review them in the app. The nutrition tracker was developed based on the Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) diet [42-44]. Participants were asked to track the foods they consumed throughout the day based on the categorical breakdown of the foods in the dietary pattern [45].

The meditation feature included categories related to mindfulness, sleep, and stress. All 3 categories had 2-, 5-, and 10-min options for the participant to choose based on their preferences and needs. For social engagement, there were dedicated lessons with suggestions on how to become more engaged within the community. Finally, participants had access to a personal health coach. The health coach was available for participants to ask questions related to cognition, cognitive decline, or behavioral health. For the purposes of this intervention, the coaches had backgrounds in personal training, nutrition, nursing, and social work.

Outcome Measures

Primary Outcome

Program engagement was the primary outcome of this pilot study. Participant engagement was evaluated based on the total sessions within the app; number of lessons read; and utilization of physical activity, nutrition, and meditation features.

Secondary Outcomes

Self-reported health was assessed through behavioral data collected in the app. These behaviors included anxiety and depression (Patient Health Questionnaire-4 [PHQ-4]), minutes of physical activity completed, nutrition scores, and sleep quantity.

Data Analysis

R 3.5.2 [46] was used to conduct all analyses. Descriptive statistics were calculated for engagement and variables related to self-reported health. *t* tests were conducted to evaluate sex differences. Engagement variables included the number of lessons completed (ie, total number of times lessons were viewed), number of times physical activity was tracked, number of times nutrition was tracked, number of times meditation was tracked, and number of times VPC test results were checked. In addition, variables related to self-reported health were defined as anxiety, depression, physical activity, nutrition, and sleep quantity. Paired *t* tests were utilized to determine differences between pre- and postscores for variables related to self-reported health status.

For relationship analyses, Pearson product-moment correlation coefficients were utilized to investigate the relationships between individual engagement variables as well as the relationships between these in-app engagement variables and lessons completed. Similar relational analyses were conducted comparing overall nutritional scores and total minutes exercised with engagement variables.

In addition, change scores were calculated for self-reported health status and related behavior variables. For these change scores, Parson product-moment correlation coefficients were also utilized to investigate relationships between baseline scores and score changes over the course of the program.

Finally, the relationships between the number of distinct lessons completed and other engagement data were evaluated through Pearson product-moment correlation coefficients. As mentioned previously, although the participants could view the lessons as many times as they wanted, there were 17 distinct lessons released throughout the course of the program (1 introductory lesson + 16 weekly informational lessons). On the basis of the observed relationships, participants were separated into groups based on the number of distinct lessons completed ($n<10$ and $n\geq10$). Paired *t* tests were used to determine the differences in engagement and self-reported health based on cutoffs for completion of distinct lessons.

Results

The participant enrollment funnel is presented in Figure 2; of the initial 716 individuals targeted to be a part of the pilot program, 559 downloaded the app and were classified as enrolled. Of those enrolled, 501/559 (89.6%) completed the initial VPC and 242/559 (43.3%) completed the final VPC. Of those completing the final VPC, 51.7% were female and 49.3% were male.

No sex differences were observed for age (females: mean 51 years, SD 7.9 years; males: mean 51 years, SD 7.1 years; *P*=.97), total number of sessions (females: mean 92, SD 79; males: mean 81, SD 69; *P*=.26), frequency of physical activity tracking (females: mean 29, SD 34; males: mean 23, SD 29; *P*=.18), total lessons read (females: mean 22, SD 13; males: mean 19, SD 10; *P*=.13), total results checked (females: mean 5, SD 4; males: mean 5, SD 5; *P*=.99), or distinct lessons loaded (females: mean 13, SD 5; males: mean 12, SD 5; *P*=.38). Significant differences were observed between sexes for frequency of nutrition tracking (females: mean 36, SD 48; males: mean 22, SD 34; *P*=.002) and frequency of meditation tracking (females: mean 7, SD 10; males: mean 4, SD 5; *P*=.008). Differences were also observed in baseline VPC scores (females: mean 86.8, SD 9.2; males: mean 84.1, SD 7.3; *P*=.02); however, there was no significant difference in VPC change scores between males and females. In addition, the total number of minutes exercised were not significantly different (females: mean 1007, SD 1316; males: mean 856, SD 1083; *P*=.32) between males and females, but there were significant differences (females: mean 7.9, SD 1.7; males: mean 7.2, SD 1.5; *P*=.03) in overall nutritional score.

Pre- and postscores for self-reported health variables are reported in Table 2. Significant differences between timepoints were observed for the number of days per week eating fish (*P*=.01); however, no other significant changes were observed.

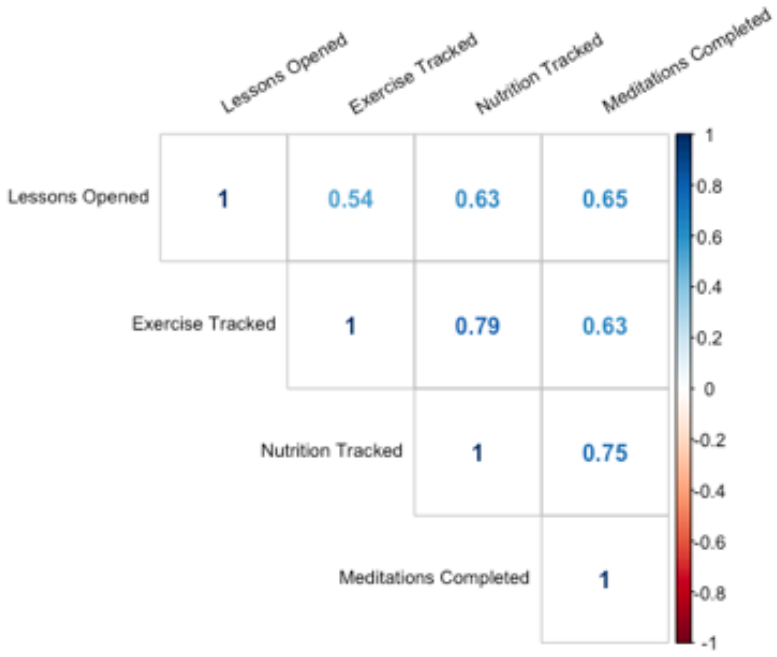
In addition, baseline VPC scores were not significantly different from final scores, indicating no cognitive changes occurred throughout the intervention.

As defined previously, the engagement variables assessed for this investigation included lessons completed, frequency of physical activity tracking, frequency of nutrition tracking, frequency of meditation tracking, and frequency of results viewing. Number of lessons completed was significantly correlated with the frequency of physical activity tracking, frequency of nutrition tracking, and frequency of meditation tracking. See Figure 3 for a visual representation of individual relationships and correlation coefficients (please note that all relationships in Figure 3 are significant).

Table 2. Analysis of pre- and postscores for Memory Health Program data.

Variable measured	Prescore, mean (SD)	Postscore, mean (SD)	*P* value
Visual paired comparison score	84.1 (8.1)	84.9 (7.2)	.09
Days per week eating fish	2.3 (1.2)	2.6 (1.2)	.01
Days per week eating vegetables	6.0 (1.5)	6.0 (1.5)	.60
Days per week sleeping more than 7 hours each night	2.8 (2.1)	3.0 (2.1)	.10
Total score (Patient Health Questionnaire-4)	2.4 (2.2)	2.6 (2.6)	.12

Figure 3. Relationships between engagement variables associated with the Memory Health Program.



When evaluating overall nutrition score, a mean final score of 7.6 (SD 1.7) was obtained by the participants (calculated based on the MIND diet recommendations [34]). Significant relationships existed between overall nutrition score (Figure 4) and frequency of nutrition tracking (*P*=.03; *r*=0.18), frequency of physical activity tracking (*P*=.02; *r*=0.19), and the total number of minutes exercised (*P*=.01; *r*=0.22). Nonsignificant correlations were observed between overall nutrition score and the total number of app sessions (*P*=.12, *r*=0.13), number of times participants meditated (*P*=.08; *r*=0.15), or total lessons read (*P*=.29; *r*=0.10). These relationships between engagement and dietary outcomes are important as scoring a 7.5 to 9.5 on the MIND diet has been associated with a 35% decreased risk of disease development (for a detailed description of the scoring structure, please see the study by Morris et al [34]).

Total minutes exercised over the course of the intervention were 936 (SD 1211) per participant. Total minutes exercised (Figure 5) was significantly correlated with total app sessions (*P*=.002; *r*=0.57), frequency of physical activity tracking (*P*<.001; *r*=0.85), frequency of nutrition tracking (*P*=.004; *r*=0.64), number of times participants meditated (*P*=.008; *r*=0.46), and total lessons read (*P*=.007; *r*=0.36). Physical activity plays a significant role in promoting cardiovascular health and reducing risk of cognitive decline [47,48]. These relationships suggest that increased Neurotrack MHP engagement supports increases in total minutes of physical activity, potentially leading to future positive health outcomes.

Figure 4. Relationship of overall nutrition score with other engagement variables.

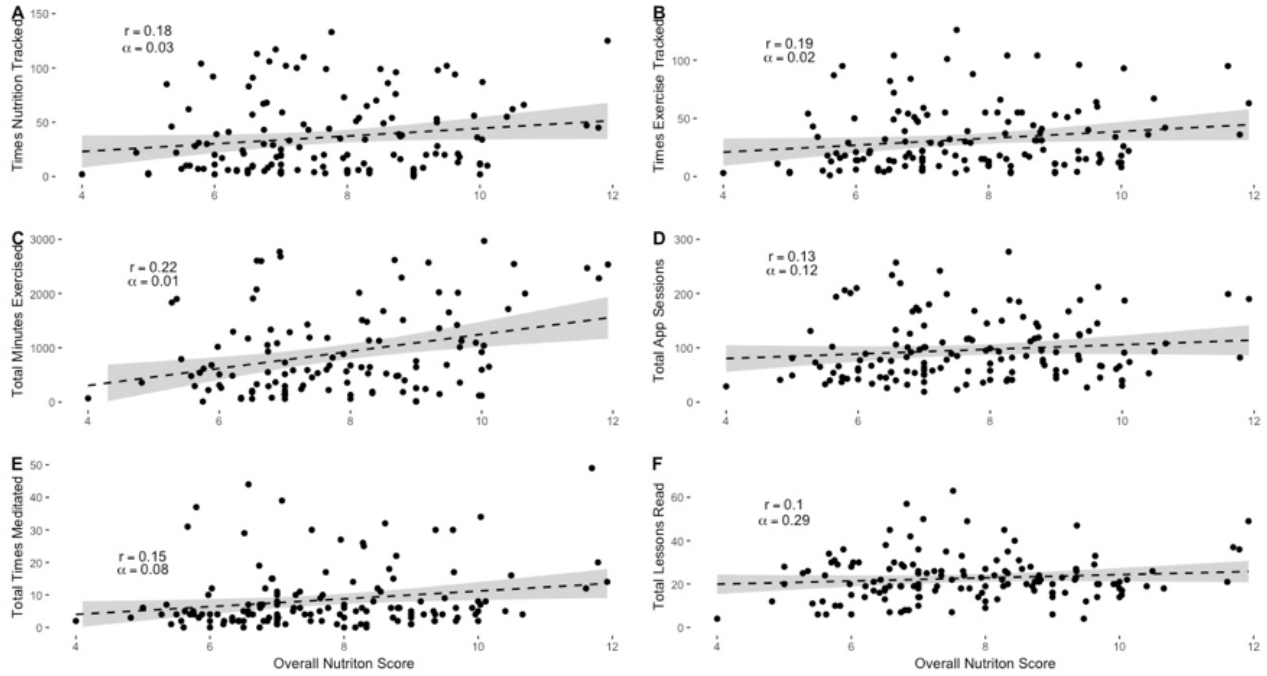
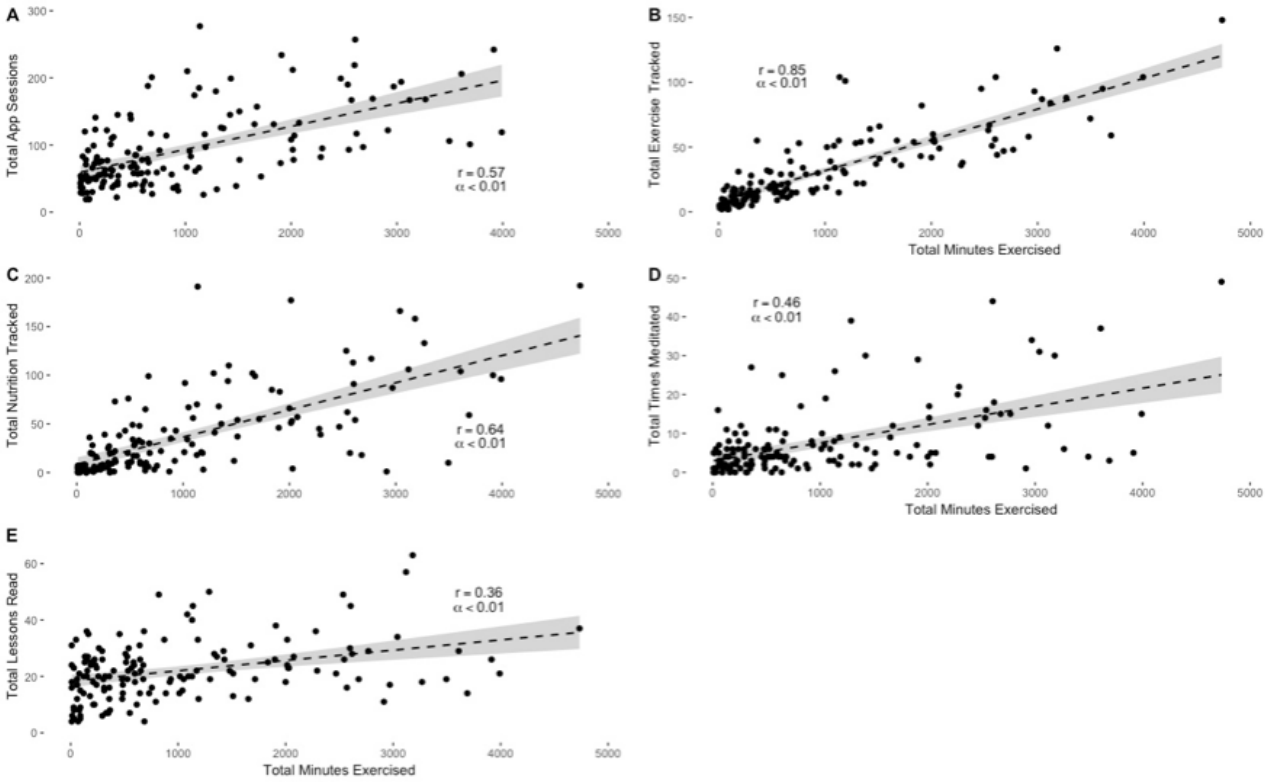


Figure 5. Total minutes exercised with other engagement variables.



Change scores were calculated for self-reported health behaviors by subtracting the prescore from the postscore. Significant negative relationships existed between self-reported baseline score and changes throughout the program. Specifically, these relationships were observed for minutes of activity per week (P=.003; r=−0.45), number of days eating fish (P=.007; r=−0.30), and days eating vegetables (P=.004; r=−0.45). These data indicate that individuals initially reporting lower levels of the behavior improved, whereas those initially reporting higher levels of the behavior remained stable.

The total number of distinct lessons completed (out of the 17 provided) were significantly related to the engagement variables (Figure 6). Specifically, distinct lessons completed were significantly related to the frequency of physical activity tracking (P=.004; r=0.40), frequency of nutrition tracking (P=.004; r=0.43), total number of times participants meditated

(*P*<.007; *r*=0.35), and total minutes exercised (*P*=.008; *r*=0.33). On the basis of these relationships, a cutoff of 10 distinct lessons completed was determined to be an inflection point for engagement. Paired *t* tests were used to determine the differences in engagement and self-reported health based on a 10-lesson cutoff point.

As shown in Table 3, no differences were observed in VPC scores (*P*=.55) between lesson completion groups (<10 lessons vs ≥10 lessons); however, significant differences were observed for total minutes exercised, frequency of physical activity tracking, frequency of nutrition tracking, and total number of times participants meditated (all *P* values <.001). When comparing change scores between lesson completion groups for self-reported health, there were no differences observed between groups for PHQ-4 score (*P*=.48), minutes of activity completed per week (*P*=.59), days per week eating fish (*P*=.93), days per week eating vegetables (*P*=.46), or nights per week sleeping more than 7 hours (*P*=.79).

Figure 6. Relationships between the total number of distinct lessons completed and other engagement variables. Note: drop lines indicate the 10-lesson cutoff.

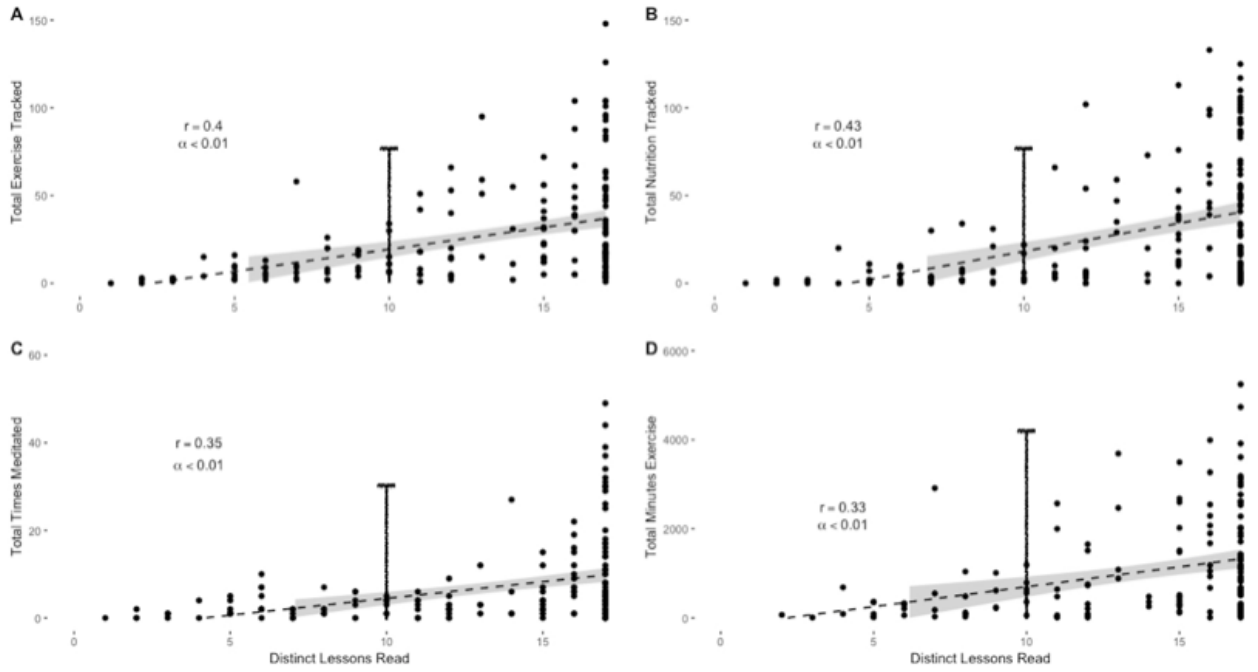


Table 3. Participant engagement data based on completion of 10 or more distinct lessons.

Engagement variable	<10 distinct lessons completed, mean (SD)	≥10 distinct lessons completed, mean (SD)	*P* value
Total minutes exercised	378 (582)	1181 (1131)	<.001
Total number of times physical activity was tracked	8.9 (10.0)	34.1 (32.3)	<.001
Total number of times nutrition was tracked	5.6 (8.7)	41.6 (46.6)	<.001
Total number of times participants meditated	2.3 (2.5)	8.9 (12.0)	<.001
Visual paired comparison score	−1.5 (5.6)	−0.9 (7.2)	.55

Discussion

Principal Findings

The primary aim of this pilot investigation was to evaluate the Neurotrack MHP's feasibility and acceptability of use in free-living Japanese adults. By combining aspects of physical activity, diet, sleep, stress, social interaction, and cognitive engagement, this new method of program delivery aimed to expand the scope and accessibility of the intervention, which are currently limiting factors in receiving proper guidance through behavior change for cognitive health [18]. The choice to utilize a Japanese population was based on the costs related to cognitive decline and disease care [12,13] of their aging population.

This investigation focused on a multidomain approach to cognitive health; multidomain lifestyle interventions are recommended as the best strategy to ameliorate cognitive decline because of the synergistic nature of multiple areas of lifestyle modification. This is similar to the original FINGER study [16] that served as the framework for this investigation. The FINGER study showed a 25% increase in global cognition, with specific increases of 83% and 150% in executive functioning and processing speed, respectively. However, these increases were not able to be mapped to a single behavioral intervention, given the multidimensional nature of the intervention. Furthermore,

the initial study evaluating the Web version of this digital intervention was unable to delineate a single factor relating to healthy improvements [28], justifying the need for an intervention that provides a more personalized approach (ie, not all participants will be able to exercise 60 min per day). In this study, we present a similar result wherein significant relationships were observed between the various types of engagement within the program (reading lessons, tracking physical activity, tracking nutrition, and meditating). Of these variables, only lessons followed a strict cadence (1 released per week), thus presenting a potential candidate for other engagement metrics to follow. All engagement metrics were interrelated, and although it is reasonable to infer that reading the weekly lessons may be a strong factor for continued engagement, it is impossible to determine a causal relationship from the current data, and further research is required to evaluate this concept.

Significant relationships were also observed between a number of the behavioral health change scores and baseline scores, indicating more improvements in participants reporting lower initial levels, whereas individuals reporting initial higher levels remained stable. Specifically, these increased change scores were observed in minutes of physical activity completed per week, number of days eating fish, and number of days eating vegetables. Although these changes are not necessarily causal to improvements in cognition, these improvements in healthy behaviors are similar to those reported in the FINGER study [16,49] where "fish intake at least twice/week," "daily intake of vegetables," and "physical activity 2 or more times/week" were increased after the 2-year intervention. In this study, these changes were independent of changes in cognition, as assessed by the VPC task; however, given the length of time required to observe improvements in cognition, as observed in previous multidomain behavior change interventions—FINGER (24 months) [16,49], MAPT (36 months) [23], Healthy Aging Through Internet Counseling in the Elderly (18 months) [50], and VC Health (12 months) [27,28]—it is likely that the intervention used in this study was too short to result in a substantial change in cognitive performance. Furthermore, this study did not limit recruitment to *at-risk* participants as was done in previous investigations [16,27,28,49], thus limiting the ability for change given the participants' baseline cognitive performance. The similar changes in behavioral health factors observed in this study as compared with the FINGER protocol [16] indicate the potential for similar improvement in cognitive performance among at-risk participants following a longer intervention duration. As a result, future investigations will need to further evaluate the long-term efficacy of the Neurotrack MHP on cognitive health outcomes. One such example of this type of investigation is the Digital Cognitive Multidomain Alzheimer's Risk Velocity [51], which is aiming to determine if this type of a broadly disseminable digital program can slow cognitive decline in at-risk participants to ultimately delay or prevent AD onset.

Physical Activity Outcomes

Additional behavioral outcomes from this investigation included the number of minutes of activity completed and nutrition scores. As expected, the total number of minutes exercised was significantly related to the frequency of physical activity tracking. The total number of minutes exercised was also related to the total app sessions, frequency of nutrition tracking, number of times participants meditated, and total lessons read. The total minutes exercised averaged almost 1000 per participant over the course of the intervention or approximately 16.5 hours of total physical activity completed over the course of the program. These data, in combination with change scores in self-reported physical activity inversely correlated with baseline levels, indicate that participants increased their activity levels throughout the program, without decreases among individuals reporting initially higher baseline values. Physical activity plays a significant role in promoting cardiovascular health and reducing risk of cognitive decline [47,48], and it has been reported that every additional hour of light-intensity physical activity is significantly associated with higher brain volumes (even among individuals not meeting current recommended guidelines) [52].

However, it should be noted that the participants in this study did not reach activity duration thresholds reported in previous investigations suggested to enhance cognition. For example, the Nurses' Health Study showed that women who walked 90 min/week demonstrated higher global cognition scores compared with women who walked less than 40 min per week [53]. Furthermore, it has been reported that individuals who are physically active for approximately 30 min per day 5 times a week experience reduced atrophy in the medial temporal lobe, a key area for memory and executive function [54]. Although the physical activity levels reached in this investigation did not amount to the recommended thresholds presented above, this pilot was concluded after 16 weeks, and further continuation may have allowed participants to continue to reach these levels. Larger-scale investigations are required to better understand the longer-term impact of the Neurotrack MHP on physical activity patterns.

Nutrition Outcomes

Nutrition has also been discussed as a primary factor in the prevention of cognitive decline [55-57]. On the basis of the MIND diet published by Morris et al [34], we scored self-reported meals throughout the investigation, with participants averaging a final score of 7.6 (SD 1.7) for their nutritional intake. Although a score of 9.6 or higher has been linked to a 53% reduction in the rate of developing AD, those scoring a 7.5 to 9.5 have been associated with a 35% decreased risk of disease development (for a detailed description of the scoring structure, please see the study by Morris et al [34]). This is important given that *perfect eating* is generally unattainable, and flexibility and fluctuation in dietary patterns is expected. In this investigation, females tracked their nutrition significantly more often than males, while also exhibiting significantly higher overall nutritional scores. Given that there is a previously established positive relationship between the act of tracking and improved health outcomes [58], it may be that simply the act of entering one's food makes individuals more aware of what they are choosing to consume, thus leading to healthier choices and improved outcomes [59,60]. It is also important to note that females are reported to be more interested in seeking out health-related information, pay more attention

to potential worldwide pandemics, and are much more attentive as to how goods purchased in everyday life affect their health than men [61].

Lesson-Related Outcomes

Although participants were able to revisit lessons as many times as they wished, there were only 17 distinct lessons that could be completed. Distinct lessons completed were significantly related to engagement variables (total number of times physical activity was tracked, total number of times nutrition was tracked, total number of times participants meditated, and total minutes exercised). Previous behavior change programs have reported that the completion of weekly lessons is significantly related to increased health outcomes measures such as weight loss [62-64] and reductions in glycosylated hemoglobin [62,63], indicating that continued use of the Neurotrack MHP may lead to additional health benefits in at-risk individuals.

Furthermore, increased lesson completion during behavior change programs has been significantly linked to subsequent improvements in nutrition and physical activity [65]. The relationship between lesson completion and improved outcomes and engagement is further supported when evaluating participants who either completed less than 10 or 10 or more lessons. Significant differences between these groups were observed for the total minutes exercised, total number of times physical activity was tracked, total number of times nutrition was tracked, and total number of times participants mediated. As a result, it is recommended that future iterations of this program aim to increase the adherence to lesson completion, as the action appears to result in subsequent improvements in other parts of the program.

Limitations

There are a few limitations associated with this study. First, given that this was a pre-post study design and there was no control arm, we were unable to compare the changes in outcomes between individuals using the Neurotrack MHP with the changes in outcomes between individuals who did not use the program. However, previous studies have reported similar results [58,62,64], suggesting that these results may remain when including a designated control, although this cannot be finalized without future investigation. In addition, as this was a pilot study, the sample size was relatively small, and larger samples are needed to make definitive conclusions. Given the demographic used for this investigation, it should also be noted that there is a high ceiling for cognitive health in this population, making it difficult to expect cognitive changes. This is not deemed an issue for this study given the nature of the pilot; however, future studies should prioritize at-risk individuals to further understand effectiveness.

There is also a potential issue related to bias in the results (ie, individuals who engaged in more healthy behaviors were, in turn, more likely to track these behaviors). Future investigations utilizing in-person research designs should work to further confirm these results. Finally, a note needs to be made about localization. This was an American program being delivered to a Japanese population, and although there was an initial layer of localization applied to the program (translation to Japanese, adjustment of wording to be more culturally applicable, etc), there are additional changes that would have made it more directly applicable to a Japanese audience. For example, berries, which are commonly included in the MIND diet, are not easily found within Japanese grocery stores and even when available, they are expensive for regular consumers. A recent version of the MIND diet was published adapting the traditional model to the Japanese market [66], which may increase the levels of tracking or program adherence. It is recommended that future studies should focus on mastering the complexity of deep localization.

Conclusions

Overall, the various engagement metrics were significantly correlated, and greater engagement was related to improved nutrition scores and increased time exercising. In addition, the relationships between distinct lessons completed and other engagement metrics suggest that there is value in focusing the program on enhancing lesson completion. This notion is further supported by the fact that individuals who completed 10 or more lessons had significantly greater program engagement than individuals who did not. Finally, females demonstrated greater levels of dietary tracking with simultaneous increases in overall nutritional scores, indicating Japanese females may be more likely to engage with behavior change programs than their male counterparts.

Acknowledgments

Sompo Holdings (Tokyo, Japan) was the sponsor of this project and covered costs associated with implementation. No other funding was used for this project.

Conflicts of Interest

None declared.

References

1. Alzheimer's Disease International. World Alzheimer Report 2018 URL: https://www.alz.co.uk/research/world-report-2018 [accessed 2005-06-19]
2. Alzheimer's Association. 2018 Alzheimer's disease facts and figures. Alzheimers Dement 2018 Mar;14(3):367-429. [doi: 10.1016/j.jalz.2018.02.001]

3. Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chételat G, Subjective Cognitive Decline Initiative (SCD-I) Working Group. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement 2014 Nov;10(6):844-852 [FREE Full text] [doi: 10.1016/j.jalz.2014.01.001] [Medline: 24798886]
4. Bäckman L, Jones S, Berger AK, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. Neuropsychology 2005 Jul;19(4):520-531. [doi: 10.1037/0894-4105.19.4.520] [Medline: 16060827]
5. Tales A, Jessen F, Butler C, Wilcock G, Phillips J, Bayer T. Subjective cognitive decline. J Alzheimers Dis 2015 Sep 24;48(Suppl 1):S1-S3 [FREE Full text] [doi: 10.3233/JAD-150719] [Medline: 26445272]
6. Jenkins A, Tales A, Tree J, Bayer A. Are we ready? The construct of subjective cognitive impairment and its utilization in clinical practice: a preliminary UK-based service evaluation. J Alzheimers Dis 2015 Sep 24;48(Suppl 1):S25-S31. [doi: 10.3233/JAD-150541] [Medline: 26445273]
7. Hugo J, Ganguli M. Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. Clin Geriatr Med 2014 Aug;30(3):421-442 [FREE Full text] [doi: 10.1016/j.cger.2014.04.001] [Medline: 25037289]
8. GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 Jan;18(1):88-106 [FREE Full text] [doi: 10.1016/S1474-4422(18)30403-4] [Medline: 30497964]
9. Alzheimer's Association. 2015 Alzheimer's disease facts and figures. Alzheimers Dement 2015 Mar;11(3):332-384. [doi: 10.1016/j.jalz.2015.02.003] [Medline: 25984581]
10. United Nations. Ageing URL: https://www.un.org/en/sections/issues-depth/ageing/ [accessed 2019-06-03]
11. United Nations. 2017. World Population Ageing 2017 URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf [accessed 2019-06-03]
12. Okamoto S. Socioeconomic factors and the risk of cognitive decline among the elderly population in Japan. Int J Geriatr Psychiatry 2019 Feb;34(2):265-271. [doi: 10.1002/gps.5015] [Medline: 30370551]
13. Sado M, Ninomiya A, Shikimoto R, Ikeda B, Baba T, Yoshimura K, et al. The estimated cost of dementia in Japan, the most aged society in the world. PLoS One 2018;13(11):e0206508 [FREE Full text] [doi: 10.1371/journal.pone.0206508] [Medline: 30418983]
14. Xu W, Tan L, Wang HF, Jiang T, Tan MS, Tan L, et al. Meta-analysis of modifiable risk factors for Alzheimer's disease. J Neurol Neurosurg Psychiatry 2015 Dec;86(12):1299-1306. [doi: 10.1136/jnnp-2015-310548] [Medline: 26294005]
15. Li JQ, Tan L, Wang HF, Tan MS, Tan L, Xu W, et al. Risk factors for predicting progression from mild cognitive impairment to Alzheimer's disease: a systematic review and meta-analysis of cohort studies. J Neurol Neurosurg Psychiatry 2016 May;87(5):476-484. [doi: 10.1136/jnnp-2014-310095] [Medline: 26001840]
16. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015 Jun 6;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]
17. The Lancet Neurology. The risks of ignoring scientific evidence. Lancet Neurol 2019 May;18(5):415. [doi: 10.1016/S1474-4422(19)30089-4] [Medline: 30981315]
18. Rao A, Manteau-Rao M, Aggarwal N. Dementia neurology deserts: what are they and where are they located in the US? Alzheimers Dement 2017;13(7):P509 [FREE Full text] [doi: 10.1016/j.jalz.2017.06.577]
19. Handels RL, Wolfs CA, Aalten P, Verhey FR, Severens JL. Determinants of care costs of patients with dementia or cognitive impairment. Alzheimer Dis Assoc Disord 2013;27(1):30-36. [doi: 10.1097/WAD.0b013e318242da1d] [Medline: 22214890]
20. National Institute on Aging. US Study to Protect Brain Health Through Lifestyle Intervention to Reduce Risk (US POINTER) URL: https://www.nia.nih.gov/alzheimers/clinical-trials/u-s-study-protect-brain-health-through-lifestyle-intervention-reduce [accessed 2019-06-13]
21. Chen C, Xu X, Chew E, Henry CJ, Koo E. Singapore intervention study to prevent cognitive impairment and disability (singer) initiative. Alzheimers Dement 2017;13(7):P1222 [FREE Full text] [doi: 10.1016/j.jalz.2017.07.408]
22. Heffernan M, Andrews G, Singh MA, Valenzuela M, Anstey KJ, Maeder AJ, Maintain Your Brain Collaborative Team, et al. Maintain your brain: protocol of a 3-year randomized controlled trial of a personalized multi-modal digital health intervention to prevent cognitive decline among community dwelling 55 to 77 year olds. J Alzheimers Dis 2019;70(s1):S221-S237 [FREE Full text] [doi: 10.3233/JAD-180572] [Medline: 30475762]
23. Carrié I, van Kan GA, Gillette-Guyonnet S, Andrieu S, Dartigues JF, Touchon J, et al. Recruitment strategies for preventive trials. The MAPT study (MultiDomain Alzheimer Preventive Trial). J Nutr Health Aging 2012 Apr;16(4):355-359. [Medline: 22499458]
24. van Charante EP, Richard E, Eurelings LS, van Dalen JW, Ligthart SA, van Bussel EF, et al. Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (preDIVA): a cluster-randomised controlled trial. Lancet 2016 Aug 20;388(10046):797-805. [doi: 10.1016/S0140-6736(16)30950-3] [Medline: 27474376]
25. World Wide Fingers. URL: http://wwfingers.com/ [accessed 2019-05-13]
26. Anstey KJ, Bahar-Fuchs A, Herath P, Kim S, Burns R, Rebok GW, et al. Body brain life: A randomized controlled trial of an online dementia risk reduction intervention in middle-aged adults at risk of Alzheimer's disease. Alzheimers Dement (N Y) 2015 Jun;1(1):72-80 [FREE Full text] [doi: 10.1016/j.trci.2015.04.003] [Medline: 29854927]

27. Bott N, Kumar S, Krebs C, Glenn JM, Madero EN, Juusola JL. A remote intervention to prevent or delay cognitive impairment in older adults: design, recruitment, and baseline characteristics of the virtual cognitive health (VC Health) study. JMIR Res Protoc 2018 Aug 13;7(8):e11368 [FREE Full text] [doi: 10.2196/11368] [Medline: 30104186]
28. Kumar S, Tran JL, Moseson H, Tai C, Glenn JM, Madero EN, et al. The impact of the virtual cognitive health program on the cognition and mental health of older adults: pre-post 12-month pilot study. JMIR Aging 2018 Nov 9;1(2):e12031 [FREE Full text] [doi: 10.2196/12031] [Medline: 31518248]
29. Coupé P, Manjón JV, Lanuza E, Catheline G. Lifespan changes of the human brain in Alzheimer's disease. Sci Rep 2019 Mar 8;9(1):3998 [FREE Full text] [doi: 10.1038/s41598-019-39809-8] [Medline: 30850617]
30. Jagust WJ, Mormino EC. Lifespan brain activity, β-amyloid, and Alzheimer's disease. Trends Cogn Sci 2011 Nov;15(11):520-526 [FREE Full text] [doi: 10.1016/j.tics.2011.09.004] [Medline: 21983147]
31. Li Y, Pan A, Wang DD, Liu X, Dhana K, Franco OH, et al. Impact of healthy lifestyle factors on life expectancies in the US population. Circulation 2018 Jul 24;138(4):345-355 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.032047] [Medline: 29712712]
32. Taylor CA, Bouldin ED, McGuire LC. Subjective cognitive decline among adults aged ≥45 years - United States, 2015-2016. MMWR Morb Mortal Wkly Rep 2018 Jul 13;67(27):753-757 [FREE Full text] [doi: 10.15585/mmwr.mm6727a1] [Medline: 30001562]
33. Law LL, Barnett F, Yau MK, Gray MA. Effects of functional tasks exercise on older adults with cognitive impairment at risk of Alzheimer's disease: a randomised controlled trial. Age Ageing 2014 Nov;43(6):813-820. [doi: 10.1093/ageing/afu055] [Medline: 24850540]
34. Morris MC, Tangney CC, Wang Y, Sacks FM, Bennett DA, Aggarwal NT. MIND diet associated with reduced incidence of Alzheimer's disease. Alzheimers Dement 2015 Sep;11(9):1007-1014 [FREE Full text] [doi: 10.1016/j.jalz.2014.11.009] [Medline: 25681666]
35. Yaffe K, Falvey CM, Hoang T. Connections between sleep and cognition in older adults. Lancet Neurol 2014 Oct;13(10):1017-1028. [doi: 10.1016/S1474-4422(14)70172-3] [Medline: 25231524]
36. Thomas AJ, O'Brien JT. Depression and cognition in older adults. Curr Opin Psychiatry 2008 Jan;21(1):8-13. [doi: 10.1097/YCO.0b013e3282f2139b] [Medline: 18281834]
37. Gulpers BJ, Voshaar RC, van Boxtel MP, Verhey FR, Köhler S. Anxiety as a risk factor for cognitive decline: a 12-year follow-up cohort study. Am J Geriatr Psychiatry 2019 Jan;27(1):42-52. [doi: 10.1016/j.jagp.2018.09.006] [Medline: 30409551]
38. Bott N, Madero EN, Glenn J, Lange A, Anderson J, Newton D, et al. Device-embedded cameras for eye tracking-based cognitive assessment: validation with paper-pencil and computerized cognitive composites. J Med Internet Res 2018 Jul 24;20(7):e11143 [FREE Full text] [doi: 10.2196/11143] [Medline: 30042093]
39. Bott NT, Lange A, Rentz D, Buffalo E, Clopton P, Zola S. Web camera based eye tracking to assess visual memory on a visual paired comparison task. Front Neurosci 2017;11:370 [FREE Full text] [doi: 10.3389/fnins.2017.00370] [Medline: 28701915]
40. Crutcher MD, Calhoun-Haney R, Manzanares CM, Lah JJ, Levey AI, Zola SM. Eye tracking during a visual paired comparison task as a predictor of early dementia. Am J Alzheimers Dis Other Demen 2009;24(3):258-266 [FREE Full text] [doi: 10.1177/1533317509332093] [Medline: 19246573]
41. Zola S, Manzanares C, Clopton P, Lah J, Levey A. Behavioral performance as a predictor of cognitive decline: detecting the transition from healthy aging to mild cognitive impairment. Alzheimers Dement 2012 Jul;8(4):S791 [FREE Full text] [doi: 10.1016/j.jalz.2013.08.276]
42. Morris MC, Tangney CC, Wang Y, Barnes LL, Bennett D, Aggarwal N. Mind diet score more predictive than DASH or Mediterranean diet scores. Alzheimers Dement 2014 Jul;10(4):P166 [FREE Full text] [doi: 10.1016/j.jalz.2014.04.164]
43. Adjibade M, Assmann KE, Julia C, Galan P, Hercberg S, Kesse-Guyot E. Prospective association between adherence to the MIND diet and subjective memory complaints in the French NutriNet-Santé cohort. J Neurol 2019 Apr;266(4):942-952. [doi: 10.1007/s00415-019-09218-y] [Medline: 30706155]
44. Koch M, Jensen MK. Association of the MIND diet with cognition and risk of Alzheimer's disease. Curr Opin Lipidol 2016 Jun;27(3):303-304. [doi: 10.1097/MOL.0000000000000304] [Medline: 27145102]
45. Marcason W. What are the components to the MIND diet? J Acad Nutr Diet 2015 Oct;115(10):1744. [doi: 10.1016/j.jand.2015.08.002] [Medline: 26407649]
46. GBIF. 2018. R: a language and environment for statistical computing URL: https://www.gbif.org/en/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2019-04-09]
47. Mandolesi L, Polverino A, Montuori S, Foti F, Ferraioli G, Sorrentino P, et al. Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. Front Psychol 2018;9:509 [FREE Full text] [doi: 10.3389/fpsyg.2018.00509] [Medline: 29755380]
48. Gomez-Pinilla F, Hillman C. The influence of exercise on cognitive abilities. Compr Physiol 2013 Jan;3(1):403-428 [FREE Full text] [doi: 10.1002/cphy.c110063] [Medline: 23720292]

49. Kivipelto M, Solomon A, Ahtiluoto S, Ngandu T, Lehtisalo J, Antikainen R, et al. The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER): study design and progress. Alzheimers Dement 2013 Nov;9(6):657-665. [doi: 10.1016/j.jalz.2012.09.012] [Medline: 23332672]
50. Richard E, Jongstra S, Soininen H, Brayne C, Moll van Charante EP, Meiller Y, et al. Healthy Ageing Through Internet Counselling in the Elderly: the HATICE randomised controlled trial for the prevention of cardiovascular disease and cognitive impairment. BMJ Open 2016 Jun 10;6(6):e010806 [FREE Full text] [doi: 10.1136/bmjopen-2015-010806] [Medline: 27288376]
51. National Institutes of Health. Research Portfolio Online Reporting Tools (RePORT) URL: https://projectreporter.nih.gov/project_info_description.cfm?aid=9777920&icde=45658631&ddparam=&ddvalue=&ddsub=&cr=2&csb=default&cs=ASC&pball= [accessed 2019-03-22]
52. Spartano NL, Davis-Plourde KL, Himali JJ, Andersson C, Pase MP, Maillard P, et al. Association of accelerometer-measured light-intensity physical activity with brain volume: the framingham heart study. JAMA Netw Open 2019 Apr 5;2(4):e192745 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2745] [Medline: 31002329]
53. Weuve J, Kang JH, Manson JE, Breteler MM, Ware JH, Grodstein F. Physical activity, including walking, and cognitive function in older women. J Am Med Assoc 2004 Sep 22;292(12):1454-1461. [doi: 10.1001/jama.292.12.1454] [Medline: 15383516]
54. Bugg JM, Head D. Exercise moderates age-related atrophy of the medial temporal lobe. Neurobiol Aging 2011 Mar;32(3):506-514 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2009.03.008] [Medline: 19386382]
55. Smith PJ, Blumenthal JA. Dietary factors and cognitive decline. J Prev Alzheimers Dis 2016 Mar;3(1):53-64 [FREE Full text] [doi: 10.14283/jpad.2015.71] [Medline: 26900574]
56. Scarmeas N, Anastasiou CA, Yannakoulia M. Nutrition and prevention of cognitive impairment. Lancet Neurol 2018 Nov;17(11):1006-1015. [doi: 10.1016/S1474-4422(18)30338-7] [Medline: 30244829]
57. Glenn JM, Madero EN, Bott NT. Dietary protein and amino acid intake: links to the maintenance of cognitive health. Nutrients 2019 Jun 12;11(6) [FREE Full text] [doi: 10.3390/nu11061315] [Medline: 31212755]
58. Toro-Ramos T, Lee DH, Kim Y, Michaelides A, Oh TJ, Kim KM, et al. Effectiveness of a Smartphone Application for the Management of Metabolic Syndrome Components Focusing on Weight Loss: A Preliminary Study. Metab Syndr Relat Disord 2017 Nov;15(9):465-473. [doi: 10.1089/met.2017.0062] [Medline: 29035677]
59. Ingels JS, Misra R, Stewart J, Lucke-Wold B, Shawley-Brzoska S. The effect of adherence to dietary tracking on weight loss: using HLM to model weight loss over time. J Diabetes Res 2017;2017:6951495 [FREE Full text] [doi: 10.1155/2017/6951495] [Medline: 28852651]
60. Hollis JF, Gullion CM, Stevens VJ, Brantley PJ, Appel LJ, Ard JD, Weight Loss Maintenance Trial Research Group. Weight loss during the intensive intervention phase of the weight-loss maintenance trial. Am J Prev Med 2008 Aug;35(2):118-126 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.013] [Medline: 18617080]
61. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. Health Promot Int 2015 Sep;30(3):736-745. [doi: 10.1093/heapro/dat063] [Medline: 23985248]
62. Sepah SC, Jiang L, Peters AL. Translating the diabetes prevention program into an online social network: validation against CDC standards. Diabetes Educ 2014 Jul;40(4):435-443. [doi: 10.1177/0145721714531339] [Medline: 24723130]
63. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital Diabetes Prevention Program: 3-year update. BMJ Open Diabetes Res Care 2017;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
64. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile Diabetes Prevention Program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
65. Wilson MG, Sweet CM, Edge MD, Madero EN, McGuire M, Pilsmaker M, et al. Evaluation of a digital behavioral counseling program for reducing risk factors for chronic disease in a workforce. J Occup Environ Med 2017 Aug;59(8):e150-e155 [FREE Full text] [doi: 10.1097/JOM.0000000000001091] [Medline: 28650899]
66. Saito S, Yamamoto Y, Ihara M. Development of a multicomponent intervention to prevent Alzheimer's disease. Front Neurol 2019;10:490 [FREE Full text] [doi: 10.3389/fneur.2019.00490] [Medline: 31139139]

Abbreviations

AD: Alzheimer disease
FINGER: Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability
MAPT: Multidomain Alzheimer's Prevention Trial
MCI: mild cognitive impairment
MIND: Mediterranean-Dietary Approaches to Stop Hypertension Intervention for Neurodegenerative Delay
MHP: Memory Health Program
PHQ-4: Patient Health Questionnaire-4

SCD: subjective cognitive decline
VC Health: Virtual Cognitive Health
VPC: visual paired comparison

Edited by G Eysenbach; submitted 01.08.19; peer-reviewed by C Li, B Hattink; comments to author 04.09.19; revised version received 16.09.19; accepted 03.10.19; published 25.10.19.

Please cite as:
Glenn J, Madero EN, Gray M, Fuseya N, Ikeda M, Kawamura T, Arita Y, Bott NT
Engagement With a Digital Platform for Multimodal Cognitive Assessment and Multidomain Intervention in a Japanese Population: Pilot, Quasi-Experimental, Longitudinal Study
JMIR Mhealth Uhealth 2019;7(10):e15733
URL: http://mhealth.jmir.org/2019/10/e15733/
doi:10.2196/15733
PMID:31654567

Original Paper

Examining the Use of Glucose and Physical Activity Self-Monitoring Technologies in Individuals at Moderate to High Risk of Developing Type 2 Diabetes: Randomized Trial

Maxine E Whelan[1,2,3], BSc, PhD; Mark W Orme[2,3,4,5], BSc, PhD; Andrew P Kingsnorth[2,3], BSc, MSc, PhD; Lauren B Sherar[2,3,6], BSc, MSc, PhD; Francesca L Denton[2,3], BSc; Dale W Esliger[2,3,6], BSc, MSc, PhD

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

[2]National Centre for Sport and Exercise Medicine, Loughborough University, Loughborough, United Kingdom

[3]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom

[4]Department of Respiratory Sciences, University of Leicester, Leicester, United Kingdom

[5]Centre for Exercise and Rehabilitation Science, National Institute for Health Research Leicester Biomedical Research Centre-Respiratory, Leicester, United Kingdom

[6]National Institute for Health Research Leicester Biomedical Research Centre-Lifestyle, Leicester, United Kingdom

Corresponding Author:
Maxine E Whelan, BSc, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 7435330992
Email: maxine.whelan@phc.ox.ac.uk

Abstract

Background: Self-monitoring of behavior (namely, diet and physical activity) and physiology (namely, glucose) has been shown to be effective in type 2 diabetes (T2D) and prediabetes prevention. By combining self-monitoring technologies, the acute physiological consequences of behaviors could be shown, prompting greater consideration to physical activity levels today, which impact the risk of developing diabetes years or decades later. However, until recently, commercially available technologies have not been able to show individuals the health benefits of being physically active.

Objective: The objective of this study was to examine the usage, feasibility, and acceptability of behavioral and physiological self-monitoring technologies in individuals at risk of developing T2D.

Methods: A total of 45 adults aged ≥40 years and at moderate to high risk of T2D were recruited to take part in a 3-arm feasibility trial. Each participant was provided with a behavioral (Fitbit Charge 2) and physiological (FreeStyle Libre flash glucose monitor) monitor for 6 weeks, masked according to group allocation. Participants were allocated to glucose feedback (4 weeks) followed by glucose and physical activity (biobehavioral) feedback (2 weeks; group 1), physical activity feedback (4 weeks) followed by biobehavioral feedback (2 weeks; group 2), or biobehavioral feedback (6 weeks; group 3). Participant usage (including time spent on the apps and number of glucose scans) was the primary outcome. Secondary outcomes were the feasibility (including recruitment and number of sensor displacements) and acceptability (including monitor wear time) of the intervention. Semistructured qualitative interviews were conducted at the 6-week follow-up appointment.

Results: For usage, time spent on the Fitbit and FreeStyle Libre apps declined over the 6 weeks for all groups. Of the FreeStyle Libre sensor scans conducted by participants, 17% (1798/10,582) recorded rising or falling trends in glucose, and 24% (13/45) of participants changed ≥1 of the physical activity goals. For feasibility, 49% (22/45) of participants completed the study using the minimum number of FreeStyle Libre sensors, and a total of 41 sensors were declared faulty or displaced. For acceptability, participants wore the Fitbit for 40.1 (SD 3.2) days, and 20% (9/45) of participants and 53% (24/45) of participants were prompted by email to charge or sync the Fitbit, respectively. Interviews unearthed participant perceptions on the study design by suggesting refinements to the eligibility criteria and highlighting important issues about the usability, wearability, and features of the technologies.

Conclusions: Individuals at risk of developing T2D engaged with wearable digital health technologies providing behavioral and physiological feedback. Modifications are required to both the study and to commercially available technologies to maximize the chances of sustained usage and behavior change. The study and intervention were feasible to conduct and acceptable to most participants.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 17545949; isrctn.com/ISRCTN17545949

(JMIR Mhealth Uhealth 2019;7(10):e14195) doi:10.2196/14195

KEYWORDS

usage; self-monitoring; feedback; behavior; physiology, wearable electronic devices; biobehavioral sciences

Introduction

The prevalence of diabetes was estimated to be more than 3 million in England in 2017 [1]. It is a fast-growing health crisis in the United Kingdom and globally [2], but given that 3 out of 5 cases of type 2 diabetes (T2D) are preventable [3], efforts to prevent the onset of T2D in people identified at increased risk are a clear public health priority.

The British Medical Association reported in 2018 the need to prioritize prevention over cure for long-term conditions to secure the long-term sustainability of the National Health Service (NHS). In parallel, the NHS Long Term Plan outlined the need to promote digitally enabled care and patient empowerment around accessing digital tools [4]. NHS Digital is a key driving force attempting to harness the power of information and technology into existing health care pathways, using technologies such as wearables and Web-based platforms. In combination, the ability to self-monitor behavior and health using wearables has created new opportunities for people to actively participate in their health care in nonclinical settings [5,6].

Flash glucose monitoring (FGM), brought onto the market in 2016, allows users to monitor their interstitial glucose fluctuations. Early studies have demonstrated reductions in hypoglycemia, increased time spent in the target range, and greater levels of patient satisfaction compared with traditional fingerstick monitoring for individuals living with type 1 diabetes and T2D [7,8]. However, this technology has not been examined in the context of diabetes prevention.

In parallel, physical activity has an important role in disease prevention [9] and is a major risk factor for long-term conditions including T2D [10]. In the laboratory setting, brief bouts of physical activity have resulted in acute reductions in postprandial glucose and insulin in normal weight, overweight, and obese adults [11-13]. With commercially available technologies increasingly capable of monitoring real-time physical activity and glucose levels, it is an exciting opportunity to see changes in glucose in relation to physical activity outside of the laboratory setting. Even if people are not diagnosed with T2D, a relationship between physical activity and glucose still exists, which means those at risk could also be targeted as a preventative strategy. This approach could offer a unique opportunity for individuals to see these relationships in a *real-world* setting, which could influence their daily behaviors and subsequently their acute health.

The Sensing Interstitial Glucose Levels to Nudge Active Lifestyles (SIGNAL) study examined the usage of FGM and physical activity self-monitoring technologies for people identified as moderate to high risk of developing T2D. Secondary objectives were to assess the feasibility and acceptability of the trial design, recruitment, methodology, and technology. Interviews with participants were also conducted to understand their perspectives of using self-monitoring technologies in the context of T2D prevention. Assessing the feasibility was crucial to evaluate whether a full trial would be feasible.

Methods

Study Design

This trial was a randomized, 3-arm feasibility trial. The full protocol has been published [14], and the study was registered prospectively (ISRCTN17545949). All participants provided written informed consent. Loughborough University's Ethics Advisory Committee provided ethical approval for the study (reference R17-P049).

Participants and Setting

Participants were recruited between May and September 2017 by circulating posters, letters, and emails across Leicestershire, United Kingdom. In brief, participants were aged ≥40 years and owned a compatible Android smartphone. Participants must not have had a self-reported diagnosis of diabetes (type 1, type 2, or gestational) or a glycated hemoglobin (HbA_{1c}) measurement of ≥6.5%. Interested individuals were directed to complete the Leicester Risk Assessment, which is a validated tool [15] to determine the level of risk for T2D via a Web-based survey (Qualtrics; Multimedia Appendix 1). Questions included age, sex, ethnic background, waist circumference, height, weight, and family history of diabetes. After completing the Web-based survey, individuals who received a score of moderate (16-24 points) or high (≥25/47 points) risk were contacted by a researcher by telephone or email and subsequently sent the participant information sheet by email if they were interested in taking part. An in-person appointment was scheduled at Loughborough University if participants continued to express an interest in taking part after reading the participant information sheet. During this in-person appointment, a point-of-care measurement of HbA_{1c} was taken to confirm eligibility before obtaining consent.

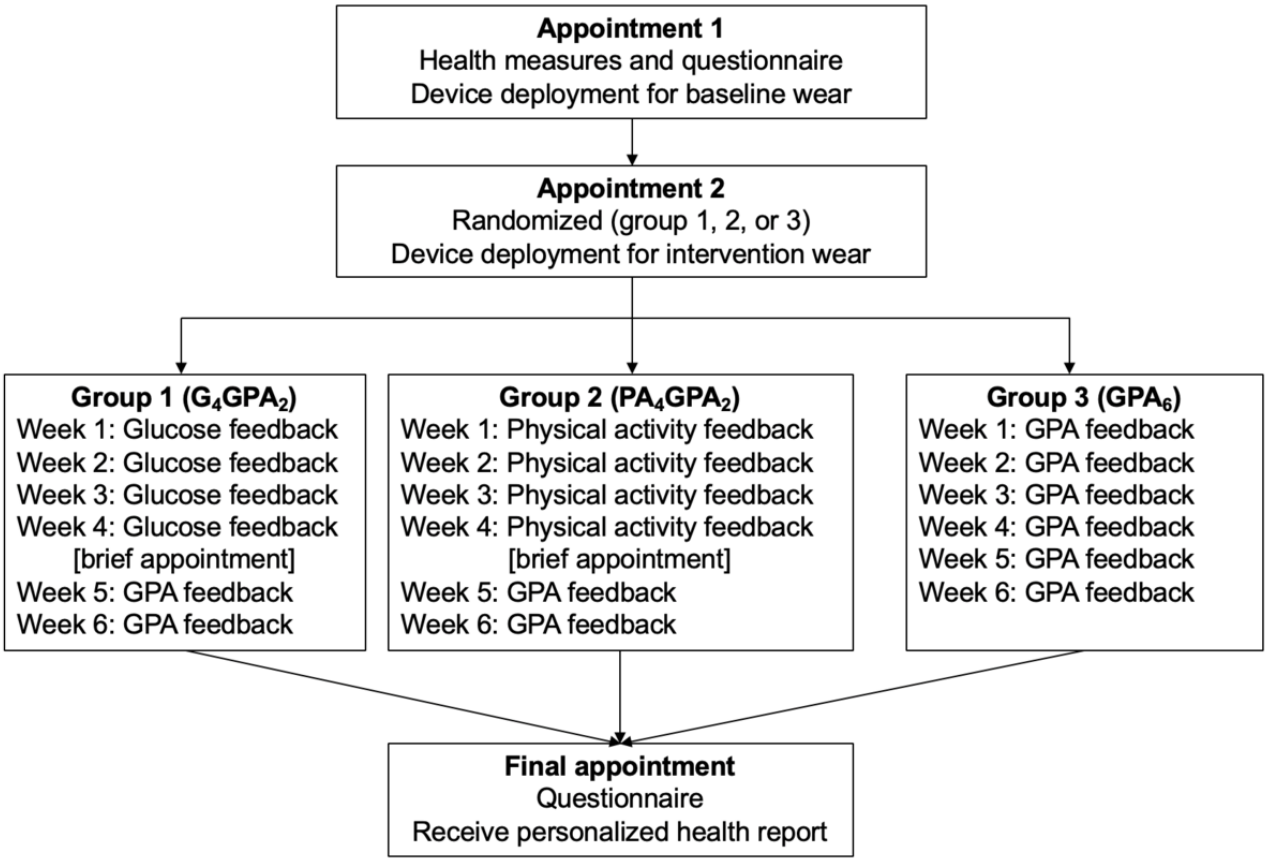
Randomization and Masking

An independent researcher produced a computer-generated randomization list with 1:1:1 allocation. After the researcher confirmed eligibility, group allocation was revealed to the participant. The researcher was informed of the participant's allocation on the day of the appointment to ensure adequate preparation (study paperwork and equipment needs varied between treatment allocations), meaning it was not possible to blind the researcher or participant to treatment allocation. Participants were encouraged to use the self-monitoring technologies as they wished, with no expectation or judgment from the researchers.

Interventions

Participants were allocated to 1 of the 3 6-week interventions (Figure 1). The 3 groups were developed so that usage could be identified for participants accessing glucose feedback alone, physical activity feedback alone, and both types of feedback (in parallel). The authors anticipated that usage would be higher in group 3 than in groups 1 and 2 from the start of the intervention. Group 1 participants were given access to glucose feedback by the FreeStyle Libre (Abbott Diabetes Care) for the first 4 weeks. In the remaining 2 weeks, these participants could access physical activity feedback by the Fitbit Charge 2 (Fitbit Inc; from here on simply referred to as Fitbit) in parallel (hereon Group 1 will be referred to as G_4GPA_2). Group 2 participants were given feedback by the Fitbit for the first 4 weeks before also accessing feedback from the FreeStyle Libre (as well as the Fitbit) for the remaining 2 weeks (hereon Group 2 will be referred to as PA_4GPA_2). Group 3 participants were given feedback from both the FreeStyle Libre and Fitbit (in parallel) for the full 6 weeks (hereon Group 3 will be referred to as GPA_6).

Figure 1. An outline of the study flow from the first appointment through to the end of participation. GPA: glucose and physical activity.



The 4+2-week design was used to identify how the patterns of use might differ depending on how the technologies were deployed. In particular, whether there were any additive benefits to receiving a single device initially before receiving feedback from the second device at a later point compared with being given access to feedback from both devices from the start, and how this affected their usage. This design was used to identify how usage varied by how the devices were deployed.

The Fitbit provided feedback regarding the number of steps taken, distance traveled, heart rate, calories expended, and flights of stairs climbed. Similarly, the FreeStyle Libre app provided feedback concerning glucose level (in mmol/L), direction of glucose trend (increasing, decreasing, or stable), time in range (above, below, or normal), and daily trends for individual days as well as previous 7, 14, or 28 days. Example screenshots of the feedback displayed by the FreeStyle Libre and Fitbit are provided in Multimedia Appendix 2. Feedback from the FreeStyle Libre was accessible via the smartphone app, and feedback from the Fitbit was accessible via the wrist-worn display and smartphone app. Both the FreeStyle Libre and Fitbit were worn throughout the 6 weeks, but settings were restricted or unrestricted, and monitors were masked or unmasked as per group allocation. If a participant should not access physical activity feedback during the study, the feedback was switched

off (via the app, hiding the feedback icons normally displayed and switching off notifications) or covered in tape (the wrist-worn display). Restricting glucose feedback meant participants wore the sensor (as normal) but did not scan it. These participants were informed that the device was automatically logging data in the background, but as they were not scanning the sensor, no data were being stored and it was considered a nonfunctioning monitor. If we allowed these participants to scan the sensor, there was no way of restricting their access to seeing feedback.

Fitbit monitors were initialized using the Fitbit app, and minute-level data were downloaded via Fitabase (Small Steps Labs LLC) and processed using Kinesoft version 3.3.80 (Kinesoft). A minimum of 3 FreeStyle Libre glucose sensors were deployed to each participant, with each offering a lifespan of 2 weeks. Glucose levels were captured by the LibreLink app and extracted in 15-min epochs using Diasend (Diasend Inc). Interstitial glucose levels were categorized as below range (<4.0 mmol/L), normal (4.0-5.9 mmol/L), or above range (>5.9 mmol/L) [16]. Scans were also characterized as rising quickly, rising, changing slowly, falling, falling quickly, or no trend arrow (determined by proprietary algorithms).

Participants were asked to ensure that the Fitbit had enough charge and was synced regularly with the Fitbit app during the intervention. Participants were notified when the battery level reached <25% or if ≥5 days had passed since a previous sync (both remotely monitored by the researchers using Fitabase). For the FreeStyle Libre, participants were asked to scan the sensor once every 7 to 8 hours to minimize data loss but were not reminded by the researchers if they failed to adhere to this. Scanning the sensor was the only way to access glucose feedback by the participants.

Procedures

We deployed ActiGraph wGT3x-BT accelerometers (ActiGraph) to measure physical activity over 7 consecutive days. These were deployed over the right hip, midclavicular line. ActiGraphs were initialized and downloaded using ActiLife (ActiGraph; Multimedia Appendix 2). Participants met the UK physical activity guidelines if they achieved a total of 150 min of moderate-to-vigorous (or ≥75 min of vigorous) physical activity in bouts of ≥10 min [17]. Participants were also asked to wear a Fitbit at baseline with settings adjusted and masked (notifications switched off and covered in tape) as to not provide feedback. They were instructed to wear them during waking hours and remove for water-based activities.

Self-reported age, sex, ethnic background, employment status, household income, highest level of education, and home postcode were recorded at baseline. Index of Multiple Deprivation was calculated using home postcode and was then segmented into 1 of the 10 categories ranging from ≤8.49 (least deprived) to ≥34.18 (most deprived) [18]. Height and waist circumference were measured. A digital scale (Tanita MC780MA) was used to measure weight and body fat. HbA_{1c} was assessed using a point-of-care Afinion AS100 Analyzer (Alere Inc), with prediabetes classified as having an HbA_{1c} of 6.0% to 6.4% [19]. Resting blood pressure was recorded using an Omron digital monitor (Omron Corporation).

From the start of the study, consecutive participants were invited to take part in a semistructured interview after completing the intervention. Interviews were completed at the 6-week follow-up appointment by a member of the research team independent from the quantitative data collection procedures and intervention delivery.

Study Outcomes

Primary Outcome

Participant usage of the Fitbit and FreeStyle Libre were assessed by time spent on the associated apps using Ethica Data (Kitchener). Usage was also assessed by the frequency with which participants scanned the FreeStyle Libre, the frequency with which the Fitbit was synced, and the number (and type) of changes to the physical activity goals. Before participants left the appointment, verbal and written information was provided about how to apply, activate, and scan the FreeStyle Libre and how to sync and charge the Fitbit. Default Fitbit physical activity goals were 10,000 steps, 30 active minutes, 10 flights of stairs, 2500 calories, and 8 kilometers per day. To record changes to these goals, the researchers checked the participant's study-specific Fitbit accounts daily via the Web-based Fitbit platform.

Secondary Outcomes

The indicators used to assess feasibility included the number of individuals who accessed and completed the Web-based survey, the number of individuals deemed eligible, uptake and retention, the number of FreeStyle Libre sensors provided to participants, and nonusage attrition [20]. Notes were made to identify the number of additional sensors provided. For acceptability, the indicators used were Fitbit wear time (defined as the presence of a heart rate signal and not categorized as sleep by Fitbit's proprietary algorithm), the number of times the research team prompted participants to sync or charge the Fitbit, the number of minutes of missing data, and the proportion of expected data for the FreeStyle Libre.

Sample Size Calculation

As typical with feasibility studies, no sample size was calculated, but a target of 45 participants was prespecified [14]. This approach was used because the trial sought to assess the feasibility of the recruitment processes.

Statistical Analyses

Descriptive statistics were reported as mean (standard deviation) or frequency (%) using Statistical Package for Social Sciences version 24.0 (SPSS Inc). Semistructured interviews were conducted at the final appointment, with a convenience sample of 26 participants, and 5 further interviews were conducted to confirm data saturation. Interviews were transcribed verbatim and analyzed thematically with the support of NVivo software, version 11 (QSR International PTY Ltd). Thematic analysis comprised data familiarization; generating initial codes; searching for, reviewing, defining, and naming themes; and producing the report [21]. Members of the research team (MO and FD) conducted initial coding. Randomly allocated subsets

of transcripts were coded by the remaining team members to ensure validity and consistency and to enhance interpretive authenticity. Team members met during data analyses to review emerging themes and to search for and collate participant views. Participants were contacted via email to provide feedback to ensure interpretations made by the team reflected the experiences of the participants [22].

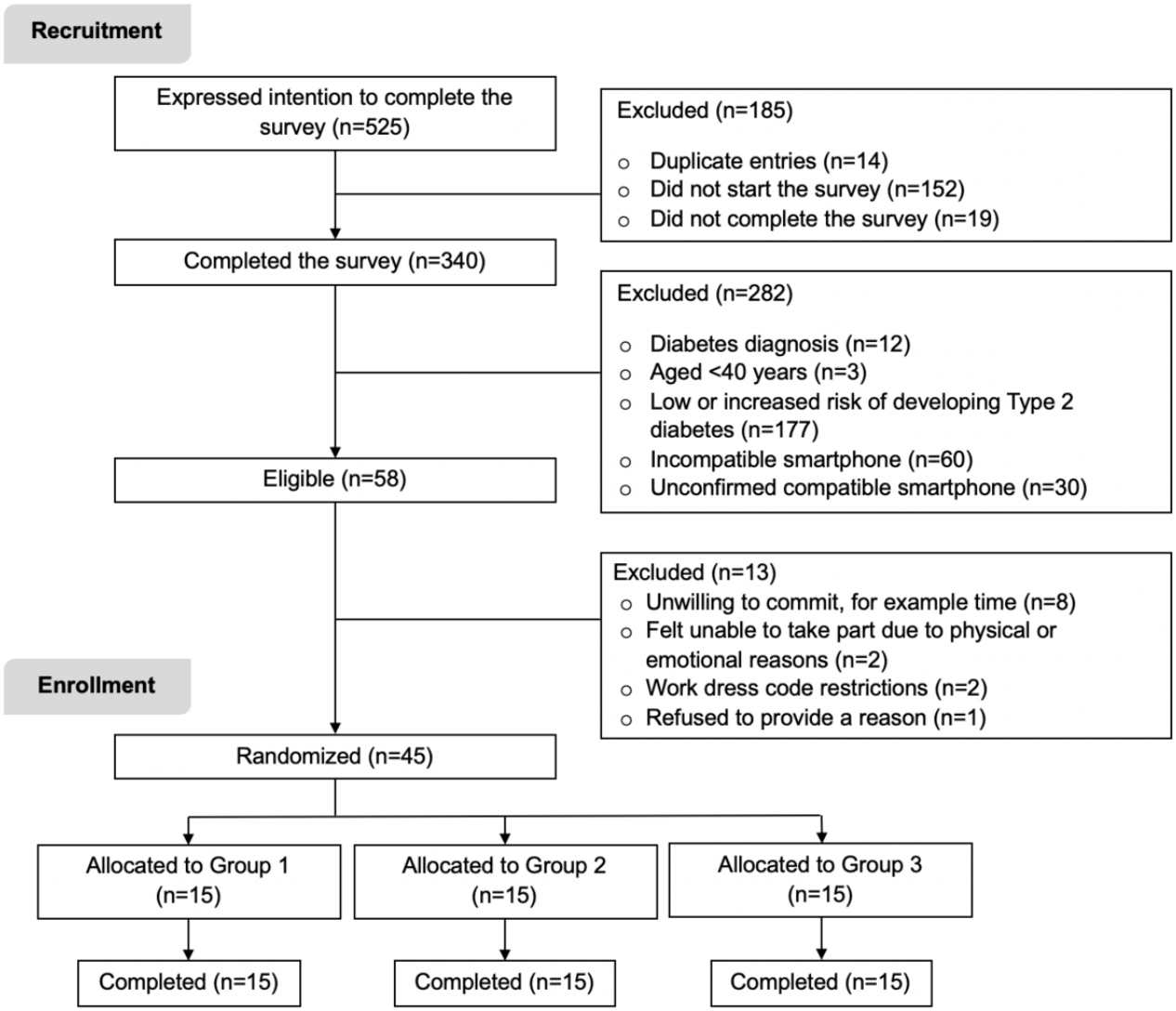
Results

Feasibility of the Trial

Eligibility, Uptake, and Retention

In total, 525 people visited the Web-based survey, 340 (64.8%) individuals completed the survey, and 58 individuals (17.1%; 11% of those visiting the survey) were eligible for the study. A total of 45 individuals (77.6% of those eligible) consented to take part, and no participants withdrew from the study (Figure 2).

Figure 2. A flow chart of participant recruitment, enrollment and allocation for the study.



Participant Characteristics

The sample was made up of more females (60%), the participants had a mean age of 56 (SD 8.7) years, and the participants were predominantly white British (88.9%) (Table 1). Most participants (53.4%) had completed undergraduate or postgraduate education, 19 (42.2%) had a household income of ≥£52,000, and 20 (44.4%) lived in a postcode considered least deprived. A total of 7 participants (15.6%) were identified as being at high risk of developing T2D, 3 (6.7%) were classified as living with prediabetes, 17 (37.8%) were overweight, and 23 (51.1%) had obesity.

The sample was highly compliant with wearing the ActiGraph and Fitbit during baseline. A total of 36 (80%) participants recorded 7 valid days of wear, with an average of 6.6 valid days and 861.5 min of daily wear recorded for the ActiGraph (Table 2). A total of 40 participants (88.9%) did not comply with the UK physical activity guidelines, and an average step count of 6905 steps was recorded.

Table 1. Participants' baseline characteristics stratified by group.

Baseline characteristics	Total sample (N=45)	Group 1: G_4GPA_2 (n=15)	Group 2: PA_4GPA_2 (n=15)	Group 3: GPA_6 (n=15)
Demographics				
Age (years), mean (SD)	56 (9)	58.8 (9)	55.3 (9)	53.9 (7)
Female gender, n (%)	27 (60)	6 (40)	9 (60)	12 (80)
Employment status, n (%)				
Employed	30 (67)	9 (60)	10 (67)	11 (73)
Retired	10 (22)	4 (27)	4 (27)	2 (13)
Other[a]	5 (11)	2 (13)	1 (7)	2 (13)
Education level, n (%)				
Postgraduate university	16 (36)	9 (60)	3 (20)	4 (27)
Undergraduate university	8 (18)	3 (20)	0 (0)	5 (33)
Some additional training	16 (36)	3 (20)	9 (60)	4 (27)
Completed secondary school	5 (11)	0 (0)	3 (20)	2 (13)
Household income (£), n (%)				
>100,000	4 (8.9)	3 (20)	1 (7)	0 (0)
52,000-100,000	15 (33)	3 (20)	4 (27)	8 (53)
18,000-51,999	18 (40)	7 (47)	7 (47)	4 (27)
<18,000	6 (13)	2 (13)	2 (13)	2 (13)
Unknown	2 (4.4)	0 (0)	1 (7)	1 (7)
Index of multiple deprivation, n (%)[b]				
Least deprived	20 (44.4)	6 (40)	7 (46.7)	7 (46.7)
Most deprived	3 (6.7)	0 (0)	1 (6.7)	2 (13.3)
Body composition, mean (SD)				
Body mass index (kg/m^2)	31.6 (7)	29.6 (5)	34.8 (9)	30.4 (4)
Waist circumference (cm)	101.5 (15)	98.8 (14)	108.4 (15)	97.4 (13)
Cardiometabolic health				
Prediabetic, n (%)	3 (7)	1 (7)	0 (0)	2 (13)
Glycated hemoglobin (measured in %), mean (SD)	5.6 (0.3)	5.6 (0.3)	5.5 (0.3)	5.6 (0.3)
Systolic BP[c] (mmHg), mean (SD)	132 (16)	135.9 (15)	131.7 (16)	128.5 (16)
Diastolic BP (mmHg), mean (SD)	81.7 (10)	82.7 (10)	79.9 (9)	82.3 (12)

[a]Other denotes looking after home and/or family, doing unpaid or voluntary work, or unable to work because of sickness or disability.

[b]Postcode deprivation offers 10 categories, but only the 2 most extreme categories have been presented for clarity.

[c]BP: blood pressure.

Table 2. Participants' baseline physical activity characteristics stratified by group.

Physical activity characteristics	Total sample (N=45)		Group 1: G_4GPA_2 (n=15)		Group 2: PA_4GPA_2 (n=15)		Group 3: GPA_6 (n=15)	
	Fitbit	ActiGraph	Fitbit	ActiGraph	Fitbit	ActiGraph	Fitbit	ActiGraph
Number of valid days, mean (SD)	6.7 (0.7)	6.6 (0.7)	6.8 (0.8)	6.8 (0.8)	6.5 (0.8)	6.5 (0.7)	6.8 (0.4)	6.5 (0.6)
Valid day, n (%); cumulative %								
7	36 (80); 80	31 (69); 69	14 (93); 93	14 (93); 93	10 (67); 67	9 (60); 60	12 (80); 80	8 (53); 53
≥6	7 (16); 96	10 (22); 91.1	0 (0); 93	0 (0); 93	4 (27); 93	4 (27); 87	3 (20); 100	6 (40); 93
≥5	0 (0); 96	3 (7); 98	0 (0); 93	0 (0); 93	0 (0); 93	2 (13); 100	0 (0); 100	1 (7); 100
≥4	2 (4); 100	1 (2); 100	1 (7); 100	1 (7); 100	1 (7); 100	0 (0); 100	0 (0); 100	0 (0); 100
Wear time (minute/day), mean (SD)	865.1 (69.6)	861.5 (86.9)	912.4 (63.3)	911.4 (88.5)	832 (61.1)	833.2 (74)	851 (60.7)	839.8 (80)
Counts per valid wear minute, mean (SD)	—[a]	328.7 (144.6)	—	342.2 (107.6)	—	281.3 (123.1)	—	362.5 (187.5)
Step count per day, mean (SD)	8575 (4530)	6905 (3776)	9329 (4251)	7331 (3433)	7650 (4007)	5637 (1963)	8747 (5367)	7748 (5148)
Sedentary (minute/day), mean (SD)	—	540.1 (95.3)	—	569.9 (90.1)	—	536.6 (89.9)	—	513.9 (103.2)
Light physical activity (minute/day), mean (SD)	—	288.2 (83.4)	—	304.4 (97.1)	—	271.4 (77.3)	—	288.7 (76.7)
MVPA[b] (minute/day), mean (SD)	—	33.1 (28.4)	—	37 (21.2)	—	25.2 (18.6)	—	37.2 (40.5)
MVPA in bouts ≥10 min (minute/day), mean (SD)	—	10.1 (21.9)	—	9.6 (14.4)	—	5 (6.2)	—	15.8 (34.5)
Met physical activity guidelines[c], n (%)	—	5 (11)	—	2 (13)	—	1 (7)	—	2 (13)
Active minutes[c,d], mean (SD)	39.8 (41.5)	—	39.5 (28.2)	—	43.5 (54.1)	—	36.9 (42.9)	—
Number of floors climbed[d], mean (SD)	11.8 (9.5)	—	13.8 (8.2)	—	7 (6.4)	—	13.8 (11.7)	—
Reminders to move[d,e,f], mean (SD)	3.6 (1.5)	—	3.3 (1.7)	—	3.9 (1.6)	—	3.6 (1.2)	—

[a]Not available from the device.

[b]MVPA: moderate-to-vigorous physical activity.

[c]Active minutes are calculated by Fitbit's undisclosed proprietary algorithm but are said to represent activity of ≥3 Metabolic Equivalent of Task (METS) and are only awarded if duration of activity is ≥10 consecutive minutes.

[d]ActiGraph accelerometers do not capture a comparable variable.

[e]Reminders to move track hourly step counts, and notifications delivered with a slight vibration are sent when the user does not reach 250 steps by 50 minutes of each hour (between 09:00 am-5:00 pm).

[f]No notifications were active during this wear period; this intends to suggest how many would have been shown had the settings been unmasked.

Feasibility of the Technology: Quantitative Insights

A total of 22 participants (48.9%) completed the study using the minimum number of 3 FreeStyle Libre sensors. Moreover, 11 of the 23 participants requiring extra sensors (47.8%) had 1 faulty or misplaced sensor, 8 (34.8%) had 2, 2 (8.7%) had 3, and the remaining 2 (8.7%) had 4. A total of 262 displacements were reported and there were more in groups G_4GPA_2 and GPA_6. It was noted that 27 participants (60%) set the LibreLink app to remind them to scan the glucose sensor. There were no instances of nonusage attrition.

Feasibility of the Technology: Qualitative Insights

Participants evaluated the technologies (a main theme) and discussed the usability, wearability, reliability, durability, preferences, privacy, and cost (as subthemes). Additional quotes to support the described subthemes are presented in Multimedia Appendix 3.

Usability

Participants generally found both technologies easy to use, in particular, how they found the Fitbit easy to charge and were complimentary about the battery life. Comments relating to applying the FreeStyle Libre emphasized the initial trepidation felt by the participants, mostly because of the visible needle and concern that the insertion process would be painful (across all 3 groups). However, there was a pleasant surprise with which the FreeStyle Libre adhered to the skin, and this nervousness typically subsided following the first application. The requirement of scanning the FreeStyle Libre at least every 8 hours to avoid data loss was highlighted as a flaw with the technology, as some participants forgot to scan regularly enough, and several participants noticed periods of missing data during sleep:

> *I thought it was all very straightforward, all very easy.* [female, GPA_6]

> *I can take it off at night and charge it at night so there's no issues with that. I put it on charge every second night really. Charging a Fitbit lasts about five days so if I put it on every second night it was fine; never had no issues with it running out at all.* [male, PA_4GPA_2]

> *I was pleasantly surprised that it didn't hurt. When I first saw the thing I thought, God this is going to be awful and hurt my arm but I couldn't believe how painless it was.* [female, GPA_6]

> *When you go to sleep, seven hours isn't all that long really. You generally need more than seven hours' sleep and not everybody would wake up in the night. I do but not everybody would, so that wasn't quite long enough.* [female, GPA_6]

Wearability

The majority of the participants found the FreeStyle Libre comfortable, with many forgetting the sensor was there (particularly participants in groups G_4GPA_2 and GPA_6). A few indicated issues with skin irritation from the Tegaderm and having trouble applying the sensor correctly. The issues with the wearability of the sensor appear to relate more to participants being self-conscious about wearing the sensor, with some resorting to covering it up with additional or alternative clothing and others suggesting refinements to the sensor and Tegaderm. Similarly, participants across all groups noted that the design of the Fitbit could be improved. In particular, participants suggested aligning the design with traditional watches, and for some, this issue meant they were not pleased when asked to replace their existing jewelry with the Fitbit. Reasons for discomfort were also related to the impact of temperature on the Fitbit strap:

> *It was fine though, I think it's brilliant. I forgot it was even there and in fact I'm a bit lost without it now it's not there, it is quite strange.* [female, GPA_6]

> *It was more about that but when I had the glucose monitor on, for a start, when it was hot, I put sleeveless on but then I was getting, what's that, what's that, what's that? So, I started wearing things with just sleeves on.* [female, PA_4GPA_2]

> *Because I have got all these lovely watches at home that I can't wear, because I have got that. Somebody at work has got one of those on, I noticed she has got it on her other hand. She has got her watch on one hand and her Fitbit on other hand. I thought well that's one way of doing it. But I couldn't wear it on there.* [female, GPA_6]

Reliability

Some participants questioned the glucose level provided by the FreeStyle Libre when it classified them as *low* or in the *red* when this was not normally the case. They tended to attribute this observation to the timing of using a new sensor or altering the location of the FreeStyle Libre on the arm. When talking about the Fitbit, the accuracy of the data provided by the technology was questioned (by the participants across all 3 groups), specifically for step count, calories, distance, mode of activity, stairs climbed, heart rate, and sleep. During masked Fitbit wear, very few problems with automatic syncing were revealed, but during unmasked periods, many participants reported the need for multiple attempts at manually syncing the Fitbit. Similarly, multiple attempts were sometimes needed to scan the FreeStyle Libre:

> *I think when I first got it, it was quite low actually for the first…well for the first day, it was below the 4 and I was thinking ooh but then it went down to the normal sort of range. And actually, the last one a week ago and the same thing actually, that's been lower, I don't know if it's the device or if I've done something different or what.* [female, G_4GPA_2]

> *I think sometimes it's been quite generous with steps. Turning around to get the towel is giving me steps and some days it said, "you've walked 'x' miles" and I thought, "I can't have." Some days it's told me I've walked eleven miles and I thought, "have I walked eleven miles?" Eleven miles is quite a…so I'm not sure it's accurate to that extent.* [male, G_4GPA_2]

> *I had a couple of issues…sometimes it's a bit difficult to link it to my phone, it just takes a bit of perseverance rather than everything happening on one flick of the screen, it might take two or three minutes just to catch on the Bluetooth. It's not a major issue, it's just a little niggle, that's probably a better word.* [male, PA_4GPA_2]

Durability

Several participants spoke about their experience of having the FreeStyle Libre sensor fall off, and, in part, this was because they were getting less aware of wearing it as time went on (more

so for participants in G_4GPA_2 and GPA_6). Particular reasons reported by participants included walking into door frames, catching the FreeStyle Libre sensor on clothing, or showers and perspiration weakening the sensor's attachment. Participants also raised annoyances about the memory of the FreeStyle Libre, suggesting that the memory is not sufficient, and perhaps, if it lasted a day or two, it would avoid some of the data losses, as some participants kept forgetting to scan it regularly:

> *I caught one on the car doorframe when I was getting something in and out of the car. The other one, I walked into a doorframe and caught it and it just bent the needle and it stopped working.* [male, GPA_6]

> *I thought it might have a better memory, last a day or two but because I kept forgetting to, because I am so busy I just kept forgetting to scan it so there are gaps and also at night time I go to bed like 9, 10 o'clock at night time and then get up sort of 12 hours later after a really good sleep and there are huge gaps.* [male, G_4GPA_2]

Privacy

Participants discussed the topic of privacy, explaining how they did not mind or have any problems with having data collected on how much they used their phone and the study apps during the study period. Some participants mentioned that this was mainly because they had nothing to hide. Others raised concerns about other apps and their monitoring activities, such as Google Maps, which captures a lot of information about user location:

> *It was interesting that you could have all this stuff going on on your phone but it also is a bit spooky as well because people are watching your performance. I suppose that, in real life, can be quite scary. It is anyway - you don't know who's watching you at any point.* [female, GPA_6]

Preference

Participants described accessing an array of feedback metrics for their activity. Preferred Fitbit features were typically step count and heart rate, with calories burned considered the least meaningful. There were conflicting remarks about the Fitbit prompts, the associated *motivational* messages were seen as childish and receiving the notification to move hourly was deemed too frequent. Nonphysical activity features such as food logging and relaxation were not used as often or considered as useful as the physical activity features of the Fitbit. This was typically because of the manual entering of data required on the app or preference for using alternative apps. The FreeStyle Libre features were talked about positively by participants with no clear preferences for the feedback provided. A suggested improvement was for greater flexibility in viewing historical data. There was an overall preference toward the FreeStyle Libre compared with the Fitbit, mostly because of the novelty of monitoring glucose levels (regardless of group):

> *I'd quite often look at my heartbeat because that was the most interesting thing on there.* [male, GPA_6]

> *It gives you those daily reports which are helpful. And I thought it was helpful that you could add in a note when you had eaten something. So, you could then align, I mean in the end you didn't really need to because you could see the patterns for yourself, but to start with that's useful that you could put in when you had eaten something. Or put a note in of some sort which is helpful.* [female, PA_4GPA_2]

> *I have actually enjoyed the glucose monitor results, looking at the graphs and charts more than the Fitbit.* [female, GPA_6]

> *I could have delved a lot more deeply into the Fitbit but I was more interested in the app and the glucose levels…I've never monitored my glucose before and I was fascinated by the whole thing. I was amazed how the device worked and all sorts of things about it.* [male, G_4GPA_2]

Cost

Several participants expressed their concern about the cost of the FreeStyle Libre sensors if they were to buy them, in particular, because they only last 2 weeks, which would restrict many people from accessing this technology:

> *It is a lot. I am sure they will come down as people use them more and more, but, and they only last two weeks, so you know it is a lot of money.* [female, PA_4GPA_2]

Acceptability of the Technology: Quantitative Insights

During the 6 weeks, 22 of 45 participants (48.9%) provided 42 days of valid Fitbit wear, with all participants averaging a total of 40.1 (SD 3.2) valid days. Compliance with syncing the Fitbit data noted that 12 participants (26.7%) received a prompt from the researchers to sync (encourage data transfer), whereas 5 (11.1%), 3 (6.7%), 2 (4.4%), and 2 (4.4%) participants received 2, 3, 4, or 5 prompts, respectively. In terms of charging the Fitbit, 9 (20%) participants received a prompt to charge the Fitbit as battery status reached <25%. No data losses were recorded for the Fitbit across the 3 groups. The level of data capture for the FreeStyle Libre was high—an average of 87.6% (SD 3.8) and 82% (SD 19) in the first and sixth week, respectively—and this was relatively consistent between the 3 groups (Multimedia Appendix 4).

There was no clear trend toward increased physical activity over the 6 weeks using step count, active minutes, number of flights of stairs, or reductions in the number of reminders to move (Multimedia Appendix 5). Similarly, there was no improvement in interstitial glucose levels using time in range (Multimedia Appendix 6).

Acceptability of the Trial: Qualitative Insights

Participants evaluated the study design (a main theme)—specifically, having a positive experience, wanting full access to feedback (in particular, participants in the groups G_4GPA_2 and PA_4GPA_2), indicating the study duration being too short, having issues with the eligibility criteria, and being uncertain of what was expected of them (as subthemes). Additional quotes to support the described subthemes are presented in Multimedia Appendix 7.

There was an overall sense of positivity in taking part in the study. Indeed, participants wanted access to the information provided by the technologies, with some explaining their frustration of having devices masked and others describing their disappointment with not being in GPA_6. The 6-week duration of the study was deemed not long enough by some participants to gain a full understanding of the relationship between their lifestyle and glucose, represent normal life, set more challenging Fitbit goals, or try out different wear locations for the FreeStyle Libre:

> *I really enjoyed it. I really learned a lot from it.* [male, G_4GPA_2]

> *Yes, I was really pleased that I was in the six-week group because I thought that will give me a lot to go on, whereas two weeks is kind of neither here nor there. I thought six weeks is a reasonable amount of time to assess things. I thought that was good for me and it led me to sort of draw some conclusions about glucose and how I was using it.* [female, GPA_6]

> *It is difficult to do that over six weeks. Perhaps over several months you might be able to pin it down to what you eat and when you eat it. Perhaps keeping a food and exercise diary and linking them together, but we weren't asked to do that.* [female, GPA_6]

Some participants suggested that the FreeStyle Libre might be better suited to individuals at greater risk of developing T2D or people already with a diagnosis of diabetes. Only including individuals with a compatible Android smartphone was highlighted as an important limitation of the study:

> *I said to [the researcher] you would have to change what app you use or something because I think you potentially would get a lot more people involved with it if it was compatible with an iPhone as well.* [female, GPA_6]

There was uncertainty on 2 fronts—changing behavior and engaging with the technology. There was some uncertainty around whether participants felt that the study required them to change their behavior or whether it was purely a monitoring or data collection exercise:

> *If I had been told, like what we are looking for is you to increase your activity because we are after this certain, we are trying to see this thing, then I would have gone for it.* [male, G_4GPA_2]

There were also inconsistencies in participants' perceptions of how much to engage with the technologies. Some individuals suggested that receiving more instructions on how to use and interpret the feedback would have been useful, whereas others were happy about not being given too much guidance and felt that the instructions given were clear. That said, participants explained how explicitly being told to be more active or to eat differently would have left them better placed to act on the feedback. Other suggestions included directing people to sources of information and being able to compare their glucose levels against a healthy profile via the FreeStyle Libre app:

> *Perhaps a bit more detailed, again just how to just sort of really make the best use of it would be better.* [female, PA_4GPA_2]

> *There's not really that much information to go with the monitor to tell you how to interpret the data.* [male, GPA_6]

> *But [the researcher] did say at the start, use this as you want it's your, your thing to use as you want. So that was good.* [male, G_4GPA_2]

Technology Usage

From weeks 1 to 6, the groups G_4GPA_2 and GPA_6 spent a lower amount of time on the FreeStyle Libre app, going from 28.3 to 12.3 min per day and 11.5 to 5.5 min per day, respectively (Table 3). PA_4GPA_2 participants logged 7.1 min per day in week 5 and 4.7 min per day in week 6. A similar pattern was observed for the FreeStyle Libre app whereby participants in groups PA_4GPA_2 and GPA_6 observed a reduction in time spent on the Fitbit app, reducing from 6.7 to 3.4 min per day and 7.6 to 3.9 min per day, respectively. Similarly, participants in G_4GPA_2 reduced their app usage from weeks 5 to 6 from 16.9 to 12.7 min per day (the only weeks when they could access it).

The average number of scans declined over time across all 3 groups (Figure 3). In the groups G_4GPA_2 and GPA_6, participants logged on average 9.4 scans per day in week 1 and 6.8 scans per day in week 6. Across weeks 5 and 6, participants in PA_4GPA_2 conducted 6.3 scans per day and 5.6 scans per day, respectively. A number of Fitbit monitors unexpectedly restored to default settings during deployment, resulting in syncs being completed automatically.

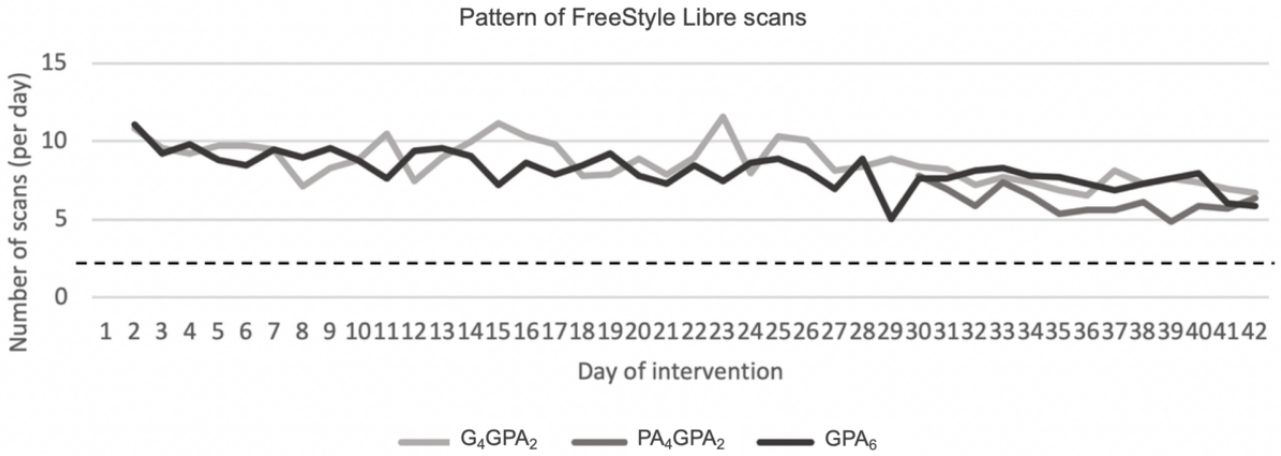
A total of 13 of 45 participants (28.9%) changed ≥1 of the physical activity goals from the default settings. Of these participants, 9 (69.2%) changed the daily step goal, whereas the number of floors, active minutes, calories, and distance goals were changed by 5 (38.5%), 3 (23.1%), 2 (15.4%), and 2 (15.4%) participants, respectively. Notably, the daily step goal was reduced by 7 participants (77.8%).

Table 3. Pattern of app usage for the Fitbit and FreeStyle Libre.

App usage	Group 1: G_4GPA_2	Group 2: PA_4GPA_2	Group 3: GPA_6
Fitbit, mean (SD)			
Week 1, minutes per day	—[a]	6.7 (3.9)	7.6 (3.8)
Week 2, minutes per day	—	5 (4.7)	4.6 (3.9)
Week 3, minutes per day	—	3.9 (2.8)	5.3 (5.0)
Week 4, minutes per day	—	3.1 (1.8)	5.5 (3.7)
Week 5, minutes per day	16.9 (16.4)	3.5 (2.1)	4.3 (3.3)
Week 6, minutes per day	12.7 (13.6)	3.4 (2.6)	3.9 (3.4)
FreeStyle Libre, mean (SD)			
Week 1, minutes per day	28.3 (23.6)	—	11.5 (8.5)
Week 2, minutes per day	21.7 (22.6)	—	7.7 (5.7)
Week 3, minutes per day	17.6 (14.7)	—	7.7 (7.9)
Week 4, minutes per day	13.2 (11.2)	—	6.3 (8.2)
Week 5, minutes per day	8.2 (5.5)	7.1 (3.9)	6.0 (5.5)
Week 6, minutes per day	12.3 (17.6)	4.7 (6.4)	5.5 (6.0)

[a]Data not available.

Figure 3. Pattern of participant scans of the glucose sensor over the 6 weeks, with the horizontal line reflecting the recommended minimum number of scans per day. One participant recorded 173 scans on day 12, deemed an outlier, and so was excluded from this figure. G_4GPA_2: Group 1; PA_4GPA_2: Group 2; GPA_6: Group 3.



Discussion

Principal Findings

To the authors' knowledge, this is the first study to deploy both behavioral and physiological self-monitoring digital health technologies to individuals at risk of developing T2D. Use of the devices reduced over the 6-week intervention period but remained higher than the minimum level of use needed to avoid data loss (eg, scanning the FreeStyle Libre every 8 hours). Despite limitations with smartphone compatibility, it was feasible to conduct the study with a high uptake rate and full retention of participants at the end of the trial. The study and technologies were acceptable to individuals at risk of T2D.

Specific Recommendations

This feasibility trial has identified several key areas for industrial and research sectors to collectively consider when considering the use of commercially available technologies in health research. These include the following: the need to (1) integrate, or facilitate the integration of, information collected by behavioral and physiological sensors into a single platform and (2) provide intelligent feedback and automated, actionable insights by using multiple data sources in real time.

Technology Usage

Participants were given minimal instruction as to how to engage with the technologies in an effort to reflect an *off the shelf* deployment. There was reduced use of the Fitbit and FreeStyle Libre over the 6-week intervention period, which was similar between the 3 groups. This suggests that the provision of both

technologies did not clearly impact the *honeymoon* period or the novelty effect frequently observed when people newly access digital health technologies [23]. For the FreeStyle Libre, the magnitude of this novelty effect was different between time spent on the app and the frequency of scanning. Scans per day reduced by approximately 28% between week 1 and week 6, whereas time spent on the app decreased by 52% to 57% for both groups accessing the FreeStyle Libre for the full 6 weeks. With participants expressing preference for more simple feedback features, these data suggest that participants ended up preferring to simply scan and see their current glucose levels rather than delve into the more detailed feedback that could be seen by the Freestyle Libre app [24,25]. This appeared to happen once users were content with the instant, and perhaps more understandable, feedback provided with each scan.

Stand-alone telemonitoring systems are an encouraging tool to support motivation [26]. The dual deployment of physical activity and physiological monitors succumbing to the issues of disuse may have been because of the feedback sources treated as distinct entities with distinct purposes. It has been noted elsewhere that relationships between behavior and health vary in how easily they can be detected by the individual [26]. This means that it may be more difficult for people to notice the benefits of being more physically active or having better blood sugar control compared with, for example, improving asthma medication adherence to relieve acute episodes of breathlessness [27].

A potentially helpful technological development would be to combine and relate the feedback provided by the Fitbit and FreeStyle Libre (or other similar devices) into a single module or feedback interface to better represent the acute relationship between behavior and physiology. With such integration, the technology may be more equipped to identify, display, and describe *teachable moments* such as "the walk you have just taken after eating bought your blood sugar back down to baseline 20 minutes faster than if you had stayed sitting down." Elsewhere, teachable moments have been defined as naturally occurring events that may motivate individuals to adopt positive, risk-reducing behaviors [28]. However, the authors propose that to display feedback showing the physiological consequences of behavior, it might need to be handled in a slightly different way to make it meaningful to the individual. Previous studies have used glucose monitoring technology to show activity-related reductions in glucose to adults living with T2D [29]. The development of such biobehavioral feedback messages will require sophisticated real-time machine learning techniques to capture and process data to produce appropriate notifications to users in an accurate and timely manner. Another provision of *teachable moments* would be to demonstrate these relationships under controlled conditions. The positive impact of physical activity bouts on acute health in the laboratory setting has been demonstrated [11,12] and can highlight key stimulus-response events. Bailey et al [30] used intermittent continuous glucose monitoring to show individuals with prediabetes and T2D their blood glucose as a tool to increase adherence to an exercise program over 8 weeks. However, this approach would not be currently feasible at the population level, which is where further technological development may still play a key role.

Feasibility of the Trial and Technology

Eligibility of individuals completing the Web-based risk survey was low (17%), with 63% of those who were ineligible for the study classified as having a low risk of developing T2D and 32% having a noncompatible smartphone. Therefore, this recruitment approach mostly attracted the *worried well* and may benefit from more targeted strategies, such as recruiting from primary care records that have identified 17.5% to 26.5% of screened individuals presenting with impaired glucose regulation [31,32]. However, since study completion, the FreeStyle Libre is now compatible with iOS devices, eliminating the limitation of smartphone compatibility that excluded many otherwise eligible survey responders. This was also highlighted by participants as a huge barrier to recruitment and a source of selection bias. That said, once recruited, all participants completed the study, describing their experience of taking part as a positive one.

Characteristics of the participants also offer important insights into the appropriateness of the recruitment strategy and eligibility criteria. Although most participants were classified as overweight or obese and did not meet physical activity recommendations, less than one-fifth of the sample was deemed at *high risk*, only 7% were prediabetic, and there was significant ethnic homogeneity, with almost 90% of the sample being white British, despite recruiting from an area with a multi-ethnic population. This is particularly important given that South Asian adults are 1.4 times more likely to develop T2D than white British adults [33]. Therefore, the next trial must actively seek more at-risk communities. After experiencing the FreeStyle Libre for at least a couple of weeks, many participants saw the glucose monitor as having great potential for people diagnosed with T2D. The FreeStyle Libre has since been made available on prescription by the NHS for people who meet a clear criterion (including currently undertake intensive monitoring >8 times daily and have an impaired awareness of hypoglycemia) with nationwide availability from April 2019 [34].

Participants expressed the need for further instruction and demonstration of the technologies and assistance with interpreting the feedback. The reported uncertainty about engaging with the technologies and hesitation in changing behavior clearly demonstrate that simply providing people with digital health technologies is insufficient for proactive lifestyle modification [35]. Although this was not the primary aim of the study, the minimal guidance approach to the study has illuminated important issues with using sophisticated digital devices in this way. The need for a human element in such prevention approaches to provide education, demonstrate devices, prescribe exercise, or motivate individuals to make positive lifestyle changes is something current digital health technologies cannot replicate or replace. In part, this may be because interfaces are not adequately intuitive or because data are not easily interpretable. Interviewees expressed difficulty knowing whether their glucose pattern was normal and what approaches they should take to improve it. Similarly, many features of the Fitbit and FreeStyle Libre went unexplored or

were regarded as surplus. Participants tended to focus on features regarded as more palatable such as step count and heart rate, which are often featured in wearable technologies [36]. Therefore, with ever more sophisticated commercially available devices, important challenges are to ensure the information provided to users is easily accessible, intuitive, and actionable.

Acceptability of the Trial and Technology

The technologies were perceived as comfortable and easy to use, and there was much positivity around participants' experience of the study, with some criticisms of the trial design linked to wanting the study to last for longer than 6 weeks as well as disappointment in not receiving the combined feedback for the full length of the intervention period (ie, being allocated to GPA_6). Adherence to wearing the Fitbit was excellent, as was charging compliance. However, issues with syncing the Fitbit data left participants frustrated and led the authors to conclude that syncing events should be interpreted with caution and were not a suitable measure of usage.

The reliability of the feedback was also bought into question, particularly the numbers presented by the Fitbit, which seemed to classify extraneous arm movements as *steps*. A comparison of baseline step count values from the Fitbit and research-grade ActiGraph may suggest an overestimation from the Fitbit. Commercially available monitors have shown strong correlations with research-grade accelerometers [37], but this does not mean numbers can be directly compared. Consequently, this perceived or real overestimation may partly explain the lack of a clear change in physical activity from baseline because of the potential for a false sense of achievement. Similar notions, although to a lesser extent, were made for the FreeStyle Libre. Some participants noticed interdevice variability and the occasional need for multiple scanning attempts, but most issues were related to the visibility of the FreeStyle Libre and the aesthetics of the device itself. The most common complaint of the FreeStyle Libre was its durability, as supported by more than half of the participants requiring at least one additional sensor. Previous studies using the FreeStyle Libre have also noted concerns with the adhesion of the sensors and have highlighted instances of mild skin irritation and bruising often linked to medical-grade adhesives [7,30]. Efforts to avoid the need for additional sensors are needed to minimize disruption for patients and costs to the NHS.

Strengths and Limitations

SIGNAL was the first study to deploy self-monitoring wearable technologies presenting behavioral and physiological feedback in real time to individuals at risk of developing T2D, classified using a validated screening questionnaire. The mixed methods design of the trial enabled real-life context to be applied to quantitative data, permitting a more in-depth assessment of participant's perspectives toward the study design and digital health technologies. The group allocations allowed all participants to experience feedback from both devices, adding to the breadth of experiences contributing to insights from the qualitative interviews. The objective measurement of physical activity allowed the sample to be compared against UK physical activity guidelines.

In addition to the limitations previously disclosed, the study had a relatively small sample size, limiting insight from statistical analyses. Assessor and participant blinding to group allocations was not possible. In addition, it was not possible to quantify what participants specifically looked at within the Fitbit and FreeStyle Libre apps. Owing to its main purpose as an intervention tool, it was not possible to set up the FreeStyle Libre in a masked mode (as achieved with the Fitbit). Consequently, no glucose data were collected from the PA_4GPA_2 group for the first 4 weeks of the intervention period. Other studies have used the FreeStyle Libre Pro model that only has a logging mode but is not currently available in all countries [38].

Conclusions

SIGNAL, to the authors' knowledge, was the first study to explore levels of use when providing combined behavioral and physiological self-monitoring wearable technologies to individuals at risk of developing T2D. This study highlights several important areas for future research, notably (1) the inclusion of a more diverse pool of individuals at risk of developing T2D or identified as living with prediabetes and (2) the detection and presentation of teachable moments linking behavioral choices with acute physiological consequences to individuals in controlled and free-living settings. Improvements to the usability, wearability, reliability, and durability are needed before such approaches to disease prevention can be implemented into routine health care. However, issues around the integration of feedback from multiple sensors and the need for real-time, actionable feedback need to be resolved for any future trials.

Acknowledgments

This study was funded in part by philanthropic support received from the late Dr the Honorable David Saul and an Engineering and Physical Sciences Research Council grant entitled "Small items of research equipment at Loughborough University" (EP/K03118x/1). This study has also been supported in part by the Higher Education Institution Challenge for Patient Supported Quality Improvement and Education in Health and Social Care (funded by the East Midlands Academic Health Science Network) for the involvement of members of the public in research and by Loughborough University School of Sport, Exercise, and Health Science for research facilitation funds.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online survey for assessing the risk of developing type 2 diabetes.
[PDF File (Adobe PDF File), 251 KB - mhealth_v7i10e14195_app1.pdf]

Multimedia Appendix 2
Application screenshots and device deployment settings.
[PDF File (Adobe PDF File), 370 KB - mhealth_v7i10e14195_app2.pdf]

Multimedia Appendix 3
All quotes pertaining to the technologies.
[PDF File (Adobe PDF File), 381 KB - mhealth_v7i10e14195_app3.pdf]

Multimedia Appendix 4
Additional Freestyle Libre data.
[PDF File (Adobe PDF File), 213 KB - mhealth_v7i10e14195_app4.pdf]

Multimedia Appendix 5
Additional Physical activity data.
[PDF File (Adobe PDF File), 299 KB - mhealth_v7i10e14195_app5.pdf]

Multimedia Appendix 6
Additional Freestyle Libre data#2.
[PDF File (Adobe PDF File), 226 KB - mhealth_v7i10e14195_app6.pdf]

Multimedia Appendix 7
All quotes pertaining to the wider study.
[PDF File (Adobe PDF File), 270 KB - mhealth_v7i10e14195_app7.pdf]

Multimedia Appendix 8
CONSORT-EHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 2545 KB - mhealth_v7i10e14195_app8.pdf]

References

1. Diabetes UK. 2017. Diabetes Prevalence 2017 URL: https://www.diabetes.org.uk/professionals/position-statements-reports/statistics/diabetes-prevalence-2017 [accessed 2019-09-13]
2. Ogurtsova K, Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF diabetes atlas: global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
3. Diabetes UK. 2018. Number of People Living With Diabetes Doubles in Twenty Years URL: https://www.diabetes.org.uk/about_us/news/diabetes-prevalence-statistics [accessed 2019-09-13]
4. National Health Services. 2019. The NHS Long Term Plan URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2019-09-13]
5. Appelboom G, Camacho E, Abraham ME, Bruce SS, Dumont EL, Zacharia BE, et al. Smart wearable body sensors for patient self-assessment and monitoring. Arch Public Health 2014;72(1):28 [FREE Full text] [doi: 10.1186/2049-3258-72-28] [Medline: 25232478]
6. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
7. Bolinder J, Antuna R, Geelhoed-Duijvestijn P, Kröger J, Weitgasser R. Novel glucose-sensing technology and hypoglycaemia in type 1 diabetes: a multicentre, non-masked, randomised controlled trial. Lancet 2016 Nov 5;388(10057):2254-2263. [doi: 10.1016/S0140-6736(16)31535-5] [Medline: 27634581]
8. Haak T, Hanaire H, Ajjan R, Hermanns N, Riveline JP, Rayman G. Flash glucose-sensing technology as a replacement for blood glucose monitoring for the management of insulin-treated type 2 diabetes: a multicenter, open-label randomized controlled trial. Diabetes Ther 2017 Feb;8(1):55-73 [FREE Full text] [doi: 10.1007/s13300-016-0223-6] [Medline: 28000140]
9. Bouchard C, Shephard RJ, Stephens T. Physical Activity, Fitness, And Health. Champaign, Illinois: Human Kinetics; 1994.

10. Camhi SM, Sisson SB, Johnson WD, Katzmarzyk PT, Tudor-Locke C. Accelerometer-determined lifestyle activity, cardiovascular disease risk factors and metabolic syndrome. Med Sci Sports Exerc 2010 May;42(5):76. [doi: 10.1249/01.MSS.0000385556.36100.3d]
11. Peddie MC, Bone JL, Rehrer NJ, Skeaff CM, Gray AR, Perry TL. Breaking prolonged sitting reduces postprandial glycemia in healthy, normal-weight adults: a randomized crossover trial. Am J Clin Nutr 2013 Aug;98(2):358-366. [doi: 10.3945/ajcn.112.051763] [Medline: 23803893]
12. Dunstan DW, Kingwell BA, Larsen R, Healy GN, Cerin E, Hamilton MT, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012 May;35(5):976-983 [FREE Full text] [doi: 10.2337/dc11-1931] [Medline: 22374636]
13. Kingsnorth AP, Whelan ME, Sanders JP, Sherar LB, Esliger DW. Using digital health technologies to understand the association between movement behaviors and interstitial glucose: exploratory analysis. JMIR Mhealth Uhealth 2018 May 3;6(5):e114 [FREE Full text] [doi: 10.2196/mhealth.9471] [Medline: 29724703]
14. Whelan ME, Kingsnorth AP, Orme MW, Sherar LB, Esliger DW. Sensing interstitial glucose to nudge active lifestyles (SIGNAL): feasibility of combining novel self-monitoring technologies for persuasive behaviour change. Br Med J Open 2017 Oct 8;7(10):e018282 [FREE Full text] [doi: 10.1136/bmjopen-2017-018282] [Medline: 28993396]
15. Gray LJ, Taub NA, Khunti K, Gardiner E, Hiles S, Webb DR, et al. The Leicester risk assessment score for detecting undiagnosed type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. Diabet Med 2010 Aug;27(8):887-895. [doi: 10.1111/j.1464-5491.2010.03037.x] [Medline: 20653746]
16. International Diabetes Federation. 2011. Management of Postmeal Glucose: Guidelines URL: https://www.idf.org/e-library/guidelines/82-management-of-postmeal-glucose.html [accessed 2019-09-13]
17. National Health Services. 2011. Physical Activity Guidelines for Adults (19–64 Years) URL: https://www.nhs.uk/Livewell/fitness/Documents/adults-19-64-years.pdf [accessed 2019-09-13]
18. GOV.UK Developer Documentation. 2015. The English Index of Multiple Deprivation (IMD) 2015 – Guidance URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/464430/English_Index_of_Multiple_Deprivation_2015_-_Guidance.pdf [accessed 2019-09-13]
19. The National Institute for Health and Care Excellence. 2012. Type 2 Diabetes: Prevention in People at High Risk URL: https://www.nice.org.uk/guidance/ph38 [accessed 2019-09-13]
20. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Boote J, Baird W, Beecroft C. Public involvement at the design stage of primary health research: a narrative review of case examples. Health Policy 2010 Apr;95(1):10-23. [doi: 10.1016/j.healthpol.2009.11.007] [Medline: 19963299]
23. Linden A, Fenn J. Michigan Ross | University of Michigan's Ross. 2003. Understanding Gartner's Hype Cycles URL: https://www.bus.umich.edu/KresgePublic/Journals/Gartner/research/115200/115274/115274.pdf [accessed 2019-09-17]
24. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
25. IQVIA. 2015. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2019-09-13]
26. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
27. Anhøj J, Nielsen L. Quantitative and qualitative usage data of an internet-based asthma monitoring tool. J Med Internet Res 2004 Sep 3;6(3):e23 [FREE Full text] [doi: 10.2196/jmir.6.3.e23] [Medline: 15471749]
28. McBride CM, Emmons KM, Lipkus IM. Understanding the potential of teachable moments: the case of smoking cessation. Health Educ Res 2003 Apr;18(2):156-170. [doi: 10.1093/her/18.2.156] [Medline: 12729175]
29. Allen NA, Fain JA, Braun B, Chipkin SR. Continuous glucose monitoring counseling improves physical activity behaviors of individuals with type 2 diabetes: a randomized clinical trial. Diabetes Res Clin Pract 2008 Jun;80(3):371-379 [FREE Full text] [doi: 10.1016/j.diabres.2008.01.006] [Medline: 18304674]
30. Bailey T, Bode BW, Christiansen MP, Klaff LJ, Alva S. The performance and usability of a factory-calibrated flash glucose monitoring system. Diabetes Technol Ther 2015 Nov;17(11):787-794 [FREE Full text] [doi: 10.1089/dia.2014.0378] [Medline: 26171659]
31. Gray LJ, Davies MJ, Hiles S, Taub NA, Webb DR, Srinivasan BT, et al. Detection of impaired glucose regulation and/or type 2 diabetes mellitus, using primary care electronic data, in a multiethnic UK community setting. Diabetologia 2012 Apr;55(4):959-966. [doi: 10.1007/s00125-011-2432-x] [Medline: 22231125]
32. Webb DR, Gray LJ, Khunti K, Srinivasan B, Taub N, Campbell S, et al. Screening for diabetes using an oral glucose tolerance test within a western multi-ethnic population identifies modifiable cardiovascular risk: the ADDITION-Leicester study. Diabetologia 2011 Sep;54(9):2237-2246. [doi: 10.1007/s00125-011-2189-2] [Medline: 21638133]

33. Khunti K, Morris DH, Weston CL, Gray LJ, Webb DR, Davies MJ. Joint prevalence of diabetes, impaired glucose regulation, cardiovascular disease risk and chronic kidney disease in South Asians and white Europeans. PLoS One 2013;8(1):e55580 [FREE Full text] [doi: 10.1371/journal.pone.0055580] [Medline: 23383233]
34. National Health Services - England. 2018. NHS to Provide Life Changing Glucose Monitors for Type 1 Diabetes Patients URL: https://www.england.nhs.uk/2018/11/nhs-to-provide-life-changing-glucose-monitors-for-type-1-diabetes-patients [accessed 2019-09-13]
35. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BioMed Central Public Health 2017 Nov 15;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]
36. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 4;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
37. 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]
38. Saboo B, Sheth SV, Joshi S, Bhandari S, Kesavadev J, Maheshwari A, et al. Use of ambulatory glucose profile for improving monitoring and management of T2DM. J Assoc Physicians India 2018 Jul;66(7):69-71. [Medline: 31325267]

Abbreviations

FGM: flash glucose monitoring
NHS: National Health Service
SIGNAL: Sensing Interstitial Glucose Levels to Nudge Active Lifestyles
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 01.04.19; peer-reviewed by L Morrison, B Nezami; comments to author 17.07.19; revised version received 15.08.19; accepted 19.08.19; published 28.10.19.

Please cite as:
Whelan ME, Orme MW, Kingsnorth AP, Sherar LB, Denton FL, Esliger DW
Examining the Use of Glucose and Physical Activity Self-Monitoring Technologies in Individuals at Moderate to High Risk of Developing Type 2 Diabetes: Randomized Trial
JMIR Mhealth Uhealth 2019;7(10):e14195
URL: http://mhealth.jmir.org/2019/10/e14195/
doi:10.2196/14195
PMID:31661077

Original Paper

Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial

Stephanie Greer[1], PhD; Danielle Ramo[1,2], PhD; Yin-Juei Chang[1], PhD; Michael Fu[1], MD; Judith Moskowitz[3], PhD; Jana Haritatos[1], PhD

[1]Hopelab, San Francisco, CA, United States

[2]Department of Psychiatry, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States

[3]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Danielle Ramo, PhD
Hopelab
100 California Street, Suite 1150
San Francisco, CA, 94111
United States
Phone: 1 415 818 8444
Email: dramo@hopelab.org

Abstract

Background: Positive psychology interventions show promise for reducing psychosocial distress associated with health adversity and have the potential to be widely disseminated to young adults through technology.

Objective: This pilot randomized controlled trial examined the feasibility of delivering positive psychology skills via the *Vivibot* chatbot and its effects on key psychosocial well-being outcomes in young adults treated for cancer.

Methods: Young adults (age 18-29 years) were recruited within 5 years of completing active cancer treatment by using the *Vivibot* chatbot on Facebook messenger. Participants were randomized to either immediate access to *Vivibot* content (experimental group) or access to only daily emotion ratings and access to full chatbot content after 4 weeks (control). Created using a human-centered design process with young adults treated for cancer, *Vivibot* content includes 4 weeks of positive psychology skills, daily emotion ratings, video, and other material produced by survivors, and periodic feedback check-ins. All participants were assessed for psychosocial well-being via online surveys at baseline and weeks 2, 4, and 8. Analyses examined chatbot engagement and open-ended feedback on likability and perceived helpfulness and compared experimental and control groups with regard to anxiety and depression symptoms and positive and negative emotion changes between baseline and 4 weeks. To verify the main effects, follow-up analyses compared changes in the main outcomes between 4 and 8 weeks in the control group once participants had access to all chatbot content.

Results: Data from 45 young adults (36 women; mean age: 25 [SD 2.9]; experimental group: n=25; control group: n=20) were analyzed. Participants in the experimental group spent an average of 74 minutes across an average of 12 active sessions chatting with *Vivibot* and rated their experience as helpful (mean 2.0/3, SD 0.72) and would recommend it to a friend (mean 6.9/10; SD 2.6). Open-ended feedback noted its nonjudgmental nature as a particular benefit of the chatbot. After 4 weeks, participants in the experimental group reported an average reduction in anxiety of 2.58 standardized t-score units, while the control group reported an increase in anxiety of 0.7 units. A mixed-effects models revealed a trend-level (P=.09) interaction between group and time, with an effect size of 0.41. Those in the experimental group also experienced greater reductions in anxiety when they engaged in more sessions (z=–1.9, P=.06). There were no significant (or trend level) effects by group on changes in depression, positive emotion, or negative emotion.

Conclusions: The chatbot format provides a useful and acceptable way of delivering positive psychology skills to young adults who have undergone cancer treatment and supports anxiety reduction. Further analysis with a larger sample size is required to confirm this pattern.

(JMIR Mhealth Uhealth 2019;7(10):e15018) doi:10.2196/15018

KEYWORDS

chatbot; positive psychology; young adult; cancer

Introduction

A total of 70,000 adolescents and young adults, aged 15-39 years, are diagnosed with cancer each year, making cancer the leading cause of disease-related death in young people in the United States [1]. In addition to this disproportionate disease burden, there are significant and unique psychosocial needs for adolescents and young adults diagnosed with cancer [2]. Adolescents and young adults are most likely to experience depression, heightened anxiety, distress, and posttraumatic stress disorder in their first 12-24 months after completing treatment, which poses significant unmet mental health needs in the period after cancer treatment [2-4].

Several studies have suggested that 20%-30% of young people treated for cancer report moderate-to-severe psychological distress lasting into adulthood [3]. Compared to their siblings, adolescents and young adults with cancer have reported poorer overall mental health [5] and are twice as likely to report clinical levels of emotional distress [6-8]. Despite this disproportionate need, there remains a lack of age-appropriate psychosocial support for adolescents and young adults after cancer treatment, which is further complicated by the difficulty reaching this geographically dispersed population [9]. There is therefore a need to address the unmet mental health burden experienced by young people who have been diagnosed with cancer in the initial years after treatment.

Advances in the field of positive psychology have shown promise in addressing the psychological distress associated with health adversity [10]. This work is built upon the importance of positive emotion in the maintenance of psychological and physical well-being [11] and evidence that positive emotion is uniquely associated with a lower risk of morbidity and mortality in healthy and chronically ill samples, independent of the effects of negative emotion [12-15]. Interventions that explicitly target positive emotion show promise for improving health outcomes in a number of chronic illnesses, including diabetes [16,17], heart disease [18,19], hypertension [20,21], depression [22,23], and HIV [24].

In the area of adolescent and young adult cancer survivorship, there is evidence that several personal resources have been shown to mitigate negative and promote positive psychosocial outcomes among young people treated for cancer [25-27]. For example, the *Promoting Resilience in Stress Management* (PRISM) intervention was designed to promote resilience coping skills in adolescents and young adults with a cancer diagnosis. A pilot trial demonstrated the feasibility of an in-person PRISM intervention with this population [28], and a clinical trial testing the efficacy of the PRISM skills intervention compared to usual care among adolescents and young adults after cancer treatment, aged 12-25 years (N=100), showed statistically significant improvements in patient‐reported resilience and cancer‐specific quality of life, as well as reduced psychological distress [29]. Secondary analyses showed additional value of benefit finding and hope among those who got PRISM with moderate-to-large effect sizes [30]. Interventions targeting resilience and positive emotion have the potential to support mental health needs among young people treated for cancer.

Depression is common among young people with cancer [31], and interventions based on increasing positive emotion have been shown to relieve symptoms of depression in a meta-analysis [23]. Psychological interventions that specifically target depression in people with cancer have shown some efficacy but tend to be underutilized by younger adults [31]. Anxiety is also common among young people with cancer, with effective treatments either not available in the settings in which adolescents and young adults with cancer receive treatment or not offered by referring clinicians. [31]. In the absence of treatments that support adolescents and young adults after cancer treatment, there is a need to fill the gap.

Digital health interventions may extend the reach and impact of positive psychology skills interventions among young people treated for cancer. Within the vast array of digital intervention platforms, fully automated conversational agents (“chatbots”) have the distinct advantage of being perceived as accessible to youth [32] and can deliver a structured set of content that can simulate the content experienced by real-life conversation (eg, with a supportive friend).

The literature on the use of chatbots to stimulate conversation about health or to actually change health behavior is emerging. For example, Bickmore et al [33] demonstrated that a carefully designed health-related conversational agent could establish a therapeutic relationship with adults attempting to increase exercise. A review of 14 chatbots in health care found that their use is still rare relative to other areas [34]. Further, most studies designed to evaluate these chatbots were quasi-experimental in design and lack outcome measures that were clear health indicators. The notable exception was the trial evaluating the *Woebot* chatbot, which found a significant reduction in depression symptoms after 2 weeks of use in college students seeking mental health support [35]. More research is needed to understand whether chatbots are associated with changes in health behaviors and emotional distress in specific populations and through varied media including social media.

Building on evidence-based interventions in the field of positive psychology and using a human-centered design approach, Hopelab (San Francisco, CA) created a chatbot called *Vivibot*, delivered over Facebook messenger, to address psychosocial needs of young people treated for cancer. The purpose of this feasibility study was to evaluate engagement and usability of the *Vivibot* chatbot. An additional goal was to evaluate the preliminary effects of positive psychology skills delivered through *Vivibot* on key psychosocial well-being outcomes in young adults treated for cancer. Outcomes were compared across two conditions: 4 weeks of *Vivibot* exposure (experimental group) or access to daily emotion ratings through Facebook Messenger with access to the full *Vivibot* chatbot after 4 weeks (control group).

Our hypotheses were as follows: (1) After 4 weeks, exposure to *Vivibot* would result in decreased depression, anxiety, and negative emotions and increased positive emotions compared to the control condition. (2) Among the treatment group, greater engagement in chatbot lessons would be associated with better outcomes.

Methods

Study Design

This was a 4-week pilot randomized controlled trial evaluating feasibility, usability, and initial efficacy of the *Vivibot* chatbot. At 4 weeks, control participants were given full access to the content in the experimental condition. An 8-week follow-up survey thus allowed for validation of the main outcome analyses in the control group.

Participants and Recruitment

Participants were young adults, aged 18-29 years, consistent with the developmental literature [36] and the field of oncology [4]; English literate; and reported having a cancer diagnosis and completing treatment for cancer within 5 years of starting the study. They also had to have access to Facebook Messenger for the study duration (either through a Facebook or Instagram account). Participants were excluded if they did not meet age or cancer diagnosis or treatment requirements or were unable to access *Vivibot* through Facebook Messenger. They were not excluded based on the type of cancer diagnosis.

Recruitment was conducted through Facebook advertising (67% of final sample), survivorship organizations (15%), and direct email after a potential participant expressed interest at a conference or event (17%). Enrollment was managed entirely through the *Vivibot* chatbot. During the recruitment period, once a user opened the chatbot interface for the first time, he/she was automatically asked four screening questions to determine eligibility. Those eligible and interested were sent an identification code and link to the study consent survey.

Study Procedure

The Ethical and Independent Review Services [37] board approved all study procedures. Following completion of the baseline assessment, participants were instructed to return to *Vivibot* and indicate that they completed the baseline survey. The chatbot then randomized users 1:1 to one of two groups: (1) immediate access to the full *Vivibot* chatbot content (experimental group) or (2) access to daily emotion ratings through Facebook Messenger with delayed full access to the full *Vivibot* chatbot only after 4 weeks (control group).

Online assessments through Qualtrics software (Salt Lake City, UT) were administered at baseline and weeks 2, 4, and 8. Participants received US $20 Amazon gift cards for completing each survey (a total of US $80 possible compensation). After 4 weeks, the control group participants were given access to all of the *Vivibot* content. After 8 weeks, participation in the study was complete, but all participants were informed that they could continue to use the chatbot product as much or as little as they desired.

Intervention Conditions

Vivibot

Vivibot is a chatbot designed to deliver prewritten and automatically delivered material to users online via a decision tree structure. Before study enrollment, users were explicitly told that they are chatting with an automated system (not a person) and periodically reminded of this throughout study participation.

Vivibot delivers a cognitive and behavioral intervention to increase positive emotion developed by Moskowitz et al (eg, [24]). The intervention was originally based on the Stress and Coping theory and the Broaden-and-Build theory of positive emotion and focused on the teaching and practice of eight positive psychological skills: noticing and acknowledging positive events, savoring positive events, gratitude, positive reappraisal, acts of kindness, mindfulness, personal strengths, and attainable goals. Rationale for inclusion of each of the positive emotion skills is provided elsewhere [38]. This core intervention was adapted for the chatbot format by creating seven conversational teaching lessons and seven practice lessons that were repeated three times to create 28 days of content. The eight skills were covered in seven lessons by combining acknowledging and savoring positive events into one lesson set.

Before launching the pilot trial, Hopelab conducted formative work through interviews and focus groups with adolescents and young adults treated for cancer to refine content for the chatbot format and inform adaptation for delivery to a young userbase with a shared experience of cancer treatment. Upon completion of the focus groups, text was tailored based on their specific suggestions, and video and other content produced by young adults who were treated for cancer was incorporated directly into the chatbot. *Vivibot* additionally incorporated six daily emotion ratings (described in the *Measures* section) and periodic check-ins on participants' satisfaction with their interactions with the chatbot. The content is outlined in Multimedia Appendix 1, and sample user experience content is in Multimedia Appendix 2.

Control

Control participants received delayed access to the full *Vivibot* chatbot. Those randomized to this group were given a message within the chatbot saying that their access to the full content would be delayed by 4 weeks. During this time, they were asked to report daily emotion ratings but received no other chatbot content. A set of six participants in the control condition encountered a technical error after the 2-week survey. This resulted in access to the full chatbot content at 2 weeks instead of the full 4 weeks. These six participants have been excluded from reported analyses to test outcomes at 4 weeks across a clean sample.

Measures

Chatbot Feasibility/Acceptability

Engagement With the Chatbot

Full conversational history with the chatbot was examined by session. An interaction was considered a *session* if there was

engagement with the bot lasting at least two user inputs within 5 minutes and a break no longer than 5 minutes. The *total number of sessions* was calculated as the count of all sessions for an individual user, and the *total interaction time* with the chatbot was defined as the total time from the start to the end of a session summed across all sessions. In addition, an *engaged session* was defined as any session that included, at minimum, a completion of the 6-item emotion rating. Identifying engaged sessions was important because some interactions deemed *sessions* included only a participant receiving a notification to check-in and then responding that they were not available to talk. These sessions therefore did not include meaningful interactions with the chatbot content. Defining engaged sessions as requiring a user to progress to daily emotion rating completion allowed for comparisons of engaged sessions in both experimental and control groups.

Chatbot Feedback

At the completion of each skill in the experimental group, users were asked to rate how helpful they found the lesson of the day on a scale from 0 ("not really") to 3 ("yes, very"). Mean ratings across all lessons were calculated for each participant (a given participant could have more than one rating). Participants in the experimental group were also given periodic opportunities to provide open-ended feedback about the chatbot program. On the seventh interaction, participants were specifically asked "How likely would you be to recommend *Vivibot* to a friend?" (rated on a scale from 0 to 10) and "Why did you give that score?" These questions were used to assess how much participants enjoyed the chatbot and what they found particularly valuable or not valuable.

Well-Being Outcomes

Anxiety and Depression Symptoms

Anxiety and depression symptoms were measured by assessments from the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative [39]. PROMIS measures are listed as "emerging measures" for further research by the American Psychiatric Association and published in the Diagnostic and Statistical Manual of Mental Disorders - fifth edition [40]. Anxiety symptoms were measured by the 4-item PROMIS Emotional Distress-Anxiety, Short Form [39]. This measure has demonstrated clinical validity in patients with chronic health conditions including cancer [41] as well as pediatric patients [42]. Depression symptoms were measured by the 4-item PROMIS Emotional Distress-Depression, Short Form [39], which captures self-reported depression symptoms, and performs similar to legacy measures (Beck Depression Inventory, Center for Epidemiological Studies Depression) in adults diagnosed with cancer [43]. Both PROMIS measures asked participants to indicate the frequency of symptoms over the past 7 days on a five-point response scale (0-4). Item scores were summed to obtain the total raw score, which was then converted to a *T* score (mean 50, SD 10). The ranges are as follows: anxiety: t=40.3-81.6 and depression: t=41.0-79.4. The cutoffs on both scales are as follows: mild, t=55; moderate, t=60; and severe, t=70 [44].

Positive and Negative Emotions

Two-week retrospective reports of positive and negative emotions were measured with the modified version of the Differential Emotions Scale, which includes additional items to measure positive emotion [45]. Positive emotions included 10 items scored from 0 ("not at all") to 4 ("most of the time"), with ranges from 0 to 40. Negative emotions included 9 items scored similarly (range: 0-36). Additionally, daily prompts for emotion ratings were triggered in the chatbot in both the experimental and control conditions. Six discrete emotions (happy, excited, content, worried, irritable/angry, and sad) were rated on a scale from 0 ("not at all") to 8 ("extremely"). Multimedia Appendix 3 presents a depiction of this measure. Responses from the three positive and three negative emotion items were each averaged to make a single positive and negative emotion score (range: 0-24) for each time period of interest (daily for daily prompts and at each 2- and 4-week surveys).

Demographics

Demographics included year of birth, gender, year of cancer diagnosis, year of treatment completion, treatment institution, basic work/school status (full-time, part-time, none), ethnicity, highest level of education, and current city of residence.

Analyses

Chatbot Engagement

Means and SDs were reported for time spent on all sessions. Within the experimental group, means and SDs for chatbot feedback were reported based on the measures described above. Open-ended feedback was evaluated, and quotes were selected to exemplify prominent themes.

Well-Being Outcomes

Well-being outcomes were assessed using a series of multilevel mixed-effects linear models for intention to treat. Multilevel models were used because they accommodate missing data and nonindependence in observations. For each construct (anxiety, depression, positive/negative emotions), a separate mixed-effects model was used to evaluate the difference in magnitude of change from baseline to week 4 follow-up assessment as a function of intervention condition (experimental vs control) and the interaction of time by condition. Each model was evaluated on the basis of statistical significance (P<.05) of the interaction term. Given the pilot nature of the trial, in cases in which the P value approached significance, effect sizes were also examined for strength and clinically meaningful differences in main outcomes across groups.

To validate results, we separately modeled changes in primary outcomes within the control condition only to examine changes in primary outcomes from the 4-week survey (when the control condition received access to the intervention) to the 8-week follow-up (which allowed for a possible 4 week intervention window, similar to that of the experimental group). This model did not include any condition or interaction term.

To examine the relationship between chatbot engagement and well-being outcomes, we used mixed-effects models to separately model each outcome as a function of the number of engaged sessions between baseline and follow-up, intervention

condition (experimental vs control), and the interaction of engaged sessions and condition.

Results

Retention

In total, 51 participants completed a baseline assessment and were randomized to a study condition (experimental group: 25; control group: 26; Figure 1). After excluding 6 control participants who were erroneously given access to chatbot content at 2 weeks instead of 4 weeks, the final analytic sample comprised 45 participants. The rate of follow-up survey completion was 73% (33/45) at 2 weeks, 73% (33/45) at 4 weeks, and 58% (26/45) at 8 weeks. We noted differences by group, with higher rates of survey completion in the control group, reaching statistical significance in week 8 (2 weeks: 80% in the control group, 68% in the experimental group, χ^2: 0.8, *P*=.37; 4 weeks: 85% in the control group, 64% in the experimental group, χ^2: 2.5, *P*=.12; 8 weeks: 75% in the control group, 44% in the experimental group, χ^2: 4.28, *P*=.04).

Participant Characteristics

Demographic information for the 45 participants included in the final analysis is presented in Table 1. Participants were mostly female (36/45, 80%), with an average age of 25 (SD 2.9; range: 19-29 years). Participants were on average 2.7 (SD 2.0) years postdiagnosis and 1.6 (SD 1.3) years postcompletion of active cancer treatment.

Figure 1. Participant recruitment and flow through the Vivibot pilot trial.

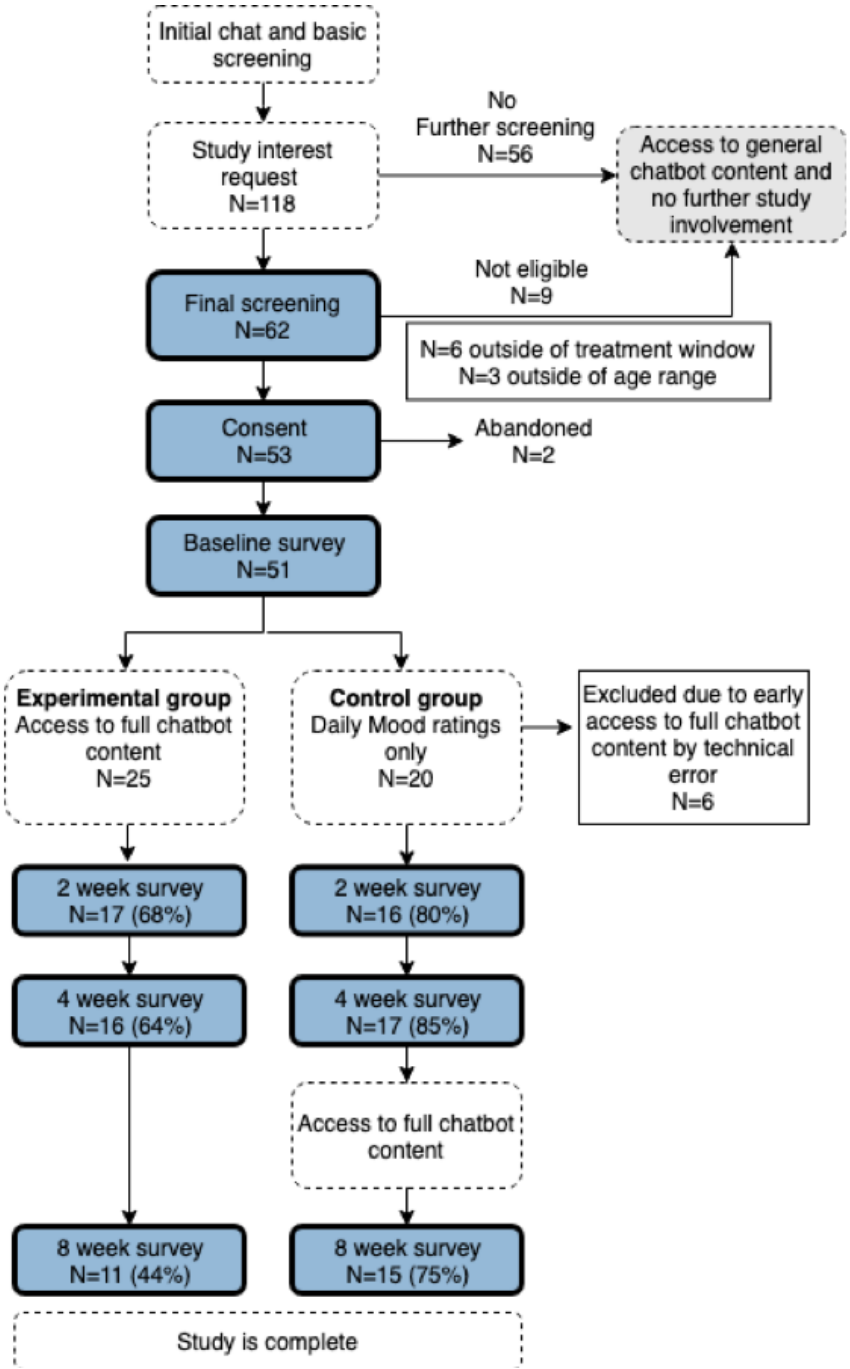


Table 1. Demographic data of participants.

Demographic	Experimental (n=25)	Control (n=20)	Total (N=45)
Age (years), mean (SD)	25 (3.1)	25 (2.9)	25 (2.9)
Gender, n (%)			
Male	4 (16)	5 (25)	9 (20)
Female	21 (84)	15 (75)	36 (80)
Education, n (%)			
Less than high school	0 (0)	2 (10)	2 (4)
High school graduate/General Educational Development	3 (12)	2 (10)	5 (11)
Some college	8 (32)	9 (45)	17 (38)
2-year college degree	2 (8)	1 (5)	3 (7)
4-year college degree	10 (40)	5 (25)	15 (33)
Master's degree	2 (8)	1 (5)	3 (7)
School/work status, n (%)			
I'm still figuring out my next move	5 (20)	3 (15)	8 (18)
I'm going to school and/or work part time	9 (36)	4 (20)	13 (29)
I'm going to school and/or work full time	11 (44)	13 (65)	24 (53)
Ethnicity, n (%)			
White/Caucasian	19 (76)	18 (90)	37 (82)
Hispanic or Latino	1 (4)	2 (10)	3 (7)
Black or African American	3 (12)	0 (0)	3 (7)
Asian or Pacific Islander	1 (4)	0 (0)	1 (2)
Prefer not to answer	1 (4)	0 (0)	1 (2)
Treatment history, mean (SD)			
Years postdiagnosis	2.7 (1.8)	2.8 (2.3)	2.7 (2.0)
Years posttreatment	1.5 (1.4)	1.8 (1.2)	1.6 (1.3)

Chatbot Engagement

During the 4 active weeks of the study, the experimental group spent an average of 73.8 (SD 52) min across an average of 12.1 (SD 7.1) engaged sessions chatting with *Vivibot*. The control group spent an average of 27.13 (SD 15.8) min across an average of 18.1 (SD 8.6) engaged sessions completing the 6-item emotion ratings only.

Perceived Helpfulness and Open-Ended Feedback.

On average, participants in the experimental group rated their experience of chatting with *Vivibot* as helpful, with an average rating of 2.03/3 (SD 0.72; range: 0-3). They were also likely to recommend *Vivibot* to a friend, with an average rating of 6.9/10 (SD 2.6; range: 0-10). When asked why they were likely or unlikely to recommend *Vivibot* to a friend, participants remarked on the utility and nonjudgmental nature of talking to an automated agent:

> *When going through treatment it was hard not to bum out my friends talking about treatments and life. My whole perspective changed. And this is a way to openly talk about those changes and you present great paths to take those thoughts rather than trying to internalize or face those awkward conversations with healthy friends.*

Additional themes related to having a shared experience with others who had undergone cancer treatment and being able to just "vent":

> *Because as weird as it is talking to a robot, it's nice to vent and be able to see others with cancer talking and speaking out about how they coped or felt during their treatment. Seeing that I'm not alone and having someone guide me to find the positives in my life now is really helpful.*

Participants also particularly enjoyed the positive psychology content itself:

> *You give me new perspectives on things and help me set goals for myself and find things to be thankful for.*
>
> *I also like the lessons you share…like looking for the good in bad situations or setting the goals to do random acts of kindness.*

Participants who gave lower ratings of *Vivibot* were less specific in their feedback:

> *I just haven't found it very helpful.*

This bot kind of makes me feel like I'm being talked at rather than talking with.

Vivbot is annoying.

Full quotes from participants can be found in Multimedia Appendix 5.

Well-Being Outcomes

Anxiety and Depression Symptoms

At baseline, both experimental and control groups presented with moderate levels of anxiety (experimental group: mean 64.5 [SD 6.1]; control group: mean 62.6 [SD 7.9]; threshold=60) and mild levels of depression (experimental group: mean 60.1 [SD 7.4]; control group: mean 59.0 [SD 9.2]; threshold=60).

Participants in the experimental group reported a greater reduction in anxiety than the control group at a trend level of statistical significance and small-to-moderate effect size (experimental reduction of 2.58 t-score units vs control of 0.7 units; z=–1.70; Cohen *d*=–0.41; *P*=.09; Table 2). Both experimental and control groups showed slight decreases in depressive symptom ratings with no evidence of a condition by time interaction (experimental reduction=1.83; control reduction=1.38; z=0.30; Cohen *d*=0.09; *P*=.77).

To validate the results, a post hoc analysis of the change between 4 weeks and 8 weeks in the control group (N=17), after receiving access to the full 28-day chatbot content, revealed a similar magnitude drop in anxiety symptoms (2.72 standardized points) and a statistical trend (*P*=.13). As in the experimental group, there was no significant reduction in depression in this subsample (Multimedia Appendix 4 presents for full results at 8 weeks).

Anxiety Symptoms by Engagement

When using engaged sessions to predict anxiety outcomes (as opposed to timepoint alone), there was a stronger trend-level relationship for group by engaged sessions interaction (z=–1.9, *P*=.06). There was no such trend for depression outcomes (z=–0.60, *P*=.55).

Positive and Negative Emotions

Both the experimental group and the control group showed a similar magnitude decrease in negative emotion in retrospective emotion ratings from baseline to week 4 (mean difference: experimental group: –0.31, control group: –0.23), with no significant or trend-level interaction by group for negative emotion at 4 weeks (z=0.23; Cohen *d*=–0.01; *P*=.97). This pattern was also reflected in the daily emotion ratings recorded in the chatbot, which showed a significant decrease in negative emotion reporting (z=–2.44; *P*=.02), but no significant interaction between groups (z=–0.74; *P*=.46).

Both the experimental group and control group showed almost no change in retrospective positive emotion ratings from baseline to week 4 (mean difference: experimental group: 0.04, control group: –0.08). Interestingly, the daily emotion ratings reported directly in the chatbot showed a significant main effect of time for increased positive emotion across groups (z=3.56; *P*=.002). Further, there was a significant interaction by group (z=–2.07; *P*=.04); however, this effect was in the opposite direction from our hypothesis, as the control group showed greater increases in positive daily emotion ratings reported directly in the chatbot, compared to the experimental group.

Table 2. Results for well-being outcomes (anxiety, depression, positive emotion, and negative emotion) across conditions and for experimental by control interactions.

Condition	Baseline, mean (SD)	Week 4, mean (SD)	Difference of means	Interaction effect size	*P* value
Anxiety				–0.41	.09
Experimental	64.5 (6.1)	61.9 (7.7)	–2.58		
Control	62.6 (7.9)	63.3 (5.5)	0.7		
Depression				0.09	.77
Experimental	60.1 (7.4)	58.2 (8.8)	–1.83		
Control	59.0 (9.2)	57.7 (6.1)	–1.38		
Negative emotion				–0.01	.97
Experimental	1.8 (0.7)	1.5 (0.9)	–0.31		
Control	1.9 (0.7)	1.6 (0.6)	–0.23		
Positive emotion				0.07	.82
Experimental	2.4 (0.8)	2.5 (1.0)	0.04		
Control	2.3 (0.9)	2.3 (0.8)	–0.08		

Discussion

Principal Findings

High engagement and positive ratings of *Vivibot* suggest that a chatbot provides a useful and acceptable format for young adults to connect with a positive psychology intervention. Qualitative responses supplemented findings from traditional metrics of user engagement such as the number of times *Vivibot* was accessed and the length of time in sessions. Feedback indicated an overall positive response to the chatbot and guided developers to generate product improvements on specific features and

content once the study was completed. Further development of this tool and potentially combining it with other person-to-person psychosocial interventions may enhance engagement even further.

Overall, positive psychology skills, delivered by a chatbot, were perceived as helpful and nonjudgmental by young adults who had undergone cancer treatment. Positive emotion, when stimulated through skills-based interventions, is thought to influence health outcomes, both directly and through the mediating influence of factors including changing health behaviors, improving physiological functioning, and increasing resources that influence health [15,46]. Possible mechanisms through which a positive psychology intervention could be influencing anxiety in young people who have undergone cancer treatment include increasing feelings of support and social control, which are influenced by positive emotion [47] and potentially poor after cancer treatment. Talking to a nonjudgmental "robot" could also have increased participants' receptivity to learn skills to manage stress, which is another core mediator of the health effects of positive emotion [15]. Additional work should be done to understand the effects of individual positive emotion intervention skills and to link skill building to subsequent long-term health impacts. The experience of a bot-based intervention as nonjudgmental shows promise for future study using this format with young people.

Based on previous work with positive psychology interventions [46], we hypothesized that engagement with *Vivibot* would result in improved psychological well-being in the form of reduced depression and anxiety symptoms and more positive and fewer negative emotions. Engagement with *Vivibot* was associated with a reduction in anxiety symptoms. Although not significant, the effect size for the main anxiety analysis of .41 in this pilot trial was small to moderate by Cohen [48] standards. Further, it was comparable to that found in a meta-analysis of psychosocial support for cancer (effect size=0.42) [49] and a positive psychology intervention for caregivers (effect size=0.33) [50]. After converting to a common scale (Multimedia Appendix 6), the change we saw in the General Anxiety Disorder-7 items equivalent scores from baseline to 4 weeks in the treatment group (reduction of 1.82 points) was greater than that found in trials testing the Woebot chatbot (reduction of 0.7 points over 2 weeks [35]) and the X2AI Artificial Intelligence tool *Tess* (reduction of 1.4 points over 4 weeks) [51]. Thus, *Vivibot* is associated with, at a minimum, a comparable reduction in anxiety symptoms as other psychosocial interventions for people with cancer and similar digital health tools available on the market for general populations.

Among those who used *Vivibot*, depression symptoms were not reduced compared to those who only completed daily emotion ratings. The active nature of the control condition in this trial could have accounted for this. The emotion rating control on its own appeared to carry some benefits for participants (rating emotions was associated with decreased negative emotion and increased positive emotion). Additionally, anecdotally, some users remarked on how completing the emotion ratings was helpful to them when asked why they would be willing to refer *Vivibot* to a friend (eg, "This is helpful to review your feelings" and "I like the daily check in questions because it reminds me to check in with myself on how I am feeling and think about why my answers might change from day to day;" Multimedia Appendix 5). This is consistent with the literature showing that emotion labeling itself can contribute to beneficial emotion regulation in response to negative events [52]. Of note, the depression symptom reduction, when converted to the Patient Health Questionnaire - 9 items equivalent, in the experimental group in our trial (1.5 points over 4 weeks) was comparable to that reported in the trial of the X2AI tool (0.9 points over 4 weeks) [51]. Since both the experimental group and the control group experienced the same emotion rating exercise, it is possible that the emotion ratings themselves carried the benefit of reductions in negative emotion, increases in positive emotion, and reductions in depression reported here in both experimental and control conditions. Further studies using a control group that did not engage in emotion labeling would be needed to more fully identify the mechanisms behind these effects.

This work extends findings showing promise for technology-enabled solutions to address mental health problems. Internet-based interventions have been shown to work at least as effectively as face-to-face intervention in improving mental health across a range of conditions. In addition to opportunities to scale interventions, online delivery offers self-pacing and the ability to incorporate practices into daily life, thereby transcending geography, time, and space [53]. Our work further suggests the chatbot interface, which is less costly than a human-delivered intervention, is a promising format to deliver psychosocial interventions to young adults. *Vivibot* included many features found to be associated with positive outcomes in internet-based interventions for depression among adolescents, including surface credibility, esthetics, and content appeal to the specific population; appropriate reduction of content compared to that in traditional interventions; and use of self-monitoring [54]. Acceptance of this positive psychology intervention is therefore not surprisingly consistent with the broad acceptance found for online-based interventions for depression and anxiety among adolescents [55].

Further, this work contributes to the growing body of literature showing promise for interventions delivered via social media in supporting short-term symptom improvement and behavior change among young adults. Most work in this area has focused on health behaviors such as smoking cessation [56] and physical activity among young people with cancer [57]; this study provides early evidence that social media–mediated intervention can be harnessed to reduce mental health symptoms. Additional randomized trials are needed to strengthen conclusions for this particular intervention and for this field in a broad sense.

Limitations

There are several limitations to this study that weaken the generalizability of these findings and motivate further investigations. First, this was a small pilot study that was not powered to detect significant effects in the psychological outcomes measured. Due to the relatively small population size and heterogeneity of treatment facilities, it is difficult to identify and recruit adolescents and young adults who had completed treatment for cancer for research studies. Our results need to be replicated in a sample large enough to reliably detect a

moderate-sized effect between groups. Second, this study was limited to participants willing and able to access *Vivibot* via Facebook Messenger. Although Facebook Messenger had the advantage of being widely accessible to this population and has almost full saturation among young adults in the United States [58], it is possible that participants did not wish to engage with Facebook Messenger in light of the controversy surrounding the privacy policies of Facebook during the study period. Other platforms, including WhatsApp and smartphone apps, should be considered in future adaptations of *Vivibot*. A third limitation was that neither the type nor stage of cancer was considered in the analyses. Due to the unique developmental and psychosocial needs of young people who have been treated for cancer, this investigation collapsed across cancer diagnoses to focus on a more targeted age range. It is possible that patients with different types of cancers would react differently to the intervention content of *Vivibot*; future work with a larger sample size should investigate this aspect further.

Conclusions

The *Vivibot* chatbot was engaging to users and resulted in anxiety reduction comparable to that found in the literature with digital health interventions targeting anxiety and depression. Given their short duration, non–artificial intelligence-based content-delivery system results, and ease with which it could be delivered in Facebook Messenger or other chat platforms, chatbots are promising for supporting the mental health needs of young people with cancer at a vulnerable time in their lives. The randomized design of this trial extends the promising findings of other positive psychology interventions delivered online [59] and gives confidence that the effect size seen with anxiety is worthy of follow-up. In addition to a larger trial, future research should consider whether the tool is as effective when delivered to other populations or platforms.

Acknowledgments

We are grateful to all of our partners who participated in the codesign of *Vivibot* and the study reported here. In particular, we would like to thank Pamela Simon and Dr Gary Dahl from the Stanford Adolescent and Young Adult Cancer Program at the Lucile Packard Children's Hospital Stanford and Stanford Health Care, and the Adolescent Young Adult Patient & Family Advisory Council for their significant input and feedback. In addition, we would like to acknowledge Ashley Pandya, Kylie Castellaw, Alexis Hoffman, Lisa Haney, Janise Chan, Julie Tinker, Robin Roskob, Tim Garcia, Phillip Lauffenburger, and Logan Daniel who were instrumental in the design and development of *Vivibot*. We also thank Drs Rebecca Block, Tckiki Davis, and Christina Armenta for helping to review and providing advice on the chatbot content.

Conflicts of Interest

The research reported here as well as the development of *Vivibot* was supported by the Hopelab Foundation. Hopelab is a nonprofit foundation that develops behavioral interventions to improve the health and well-being of young people. The design, conduct, analysis, and reporting of this study represents a scientific collaboration between Hopelab and author JM. Authors SG, DR, Y-JC, MF, and JH are employed by the Hopelab Foundation. The study sponsor was involved in the study design, collection, analysis and interpretation of data, writing of the article, and decision to submit it for publication. None of the authors have any financial or other conflicting interest in the scientific results of this study.

Editorial notice: This randomized study was not prospectively registered. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
Vivibot content overview.
[PDF File (Adobe PDF File), 210 KB - mhealth_v7i10e15018_app1.pdf]

Multimedia Appendix 2
Example user experience of positive psychology skill.
[PDF File (Adobe PDF File), 187 KB - mhealth_v7i10e15018_app2.pdf]

Multimedia Appendix 3
Screen shot of daily mood ratings.
[PDF File (Adobe PDF File), 49 KB - mhealth_v7i10e15018_app3.pdf]

Multimedia Appendix 4
Comparisons for all outcomes at 8 weeks.
[PDF File (Adobe PDF File), 43 KB - mhealth_v7i10e15018_app4.pdf]

Multimedia Appendix 5
Recommendation ratings.
[PDF File (Adobe PDF File), 98 KB - mhealth_v7i10e15018_app5.pdf]

Multimedia Appendix 6
Converted scores on the PROMIS-Anxiety and PROMIS-Depression scales.
[PDF File (Adobe PDF File), 35 KB - mhealth_v7i10e15018_app6.pdf]

Multimedia Appendix 7
CONSORT - EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3114 KB - mhealth_v7i10e15018_app7.pdf]

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018 Jan;68(1):7-30 [FREE Full text] [doi: 10.3322/caac.21442] [Medline: 29313949]
2. Kwak M, Zebrack BJ, Meeske KA, Embry L, Aguilar C, Block R, et al. Trajectories of psychological distress in adolescent and young adult patients with cancer: a 1-year longitudinal study. J Clin Oncol 2013 Jun 10;31(17):2160-2166. [doi: 10.1200/JCO.2012.45.9222] [Medline: 23650425]
3. Michel G, Rebholz CE, von der Weid NX, Bergstraesser E, Kuehni CE. Psychological distress in adult survivors of childhood cancer: the Swiss Childhood Cancer Survivor study. J Clin Oncol 2010 Apr 01;28(10):1740-1748. [doi: 10.1200/JCO.2009.23.4534] [Medline: 20194864]
4. National Cancer Institute. National Cancer Institute. 2018 Jan 31. Adolescents and Young Adults with Cancer URL: https://www.cancer.gov/types/aya [accessed 2019-02-16]
5. Phillips-Salimi CR, Lommel K, Andrykowski MA. Physical and mental health status and health behaviors of childhood cancer survivors: findings from the 2009 BRFSS survey. Pediatr Blood Cancer 2012 Jun;58(6):964-970 [FREE Full text] [doi: 10.1002/pbc.23359] [Medline: 22012636]
6. Reulen RC, Winter DL, Lancashire ER, Zeegers MP, Jenney ME, Walters SJ, et al. Health-status of adult survivors of childhood cancer: a large-scale population-based study from the British Childhood Cancer Survivor Study. Int J Cancer 2007 Aug 01;121(3):633-640 [FREE Full text] [doi: 10.1002/ijc.22658] [Medline: 17405119]
7. Zeltzer LK, Recklitis C, Buchbinder D, Zebrack B, Casillas J, Tsao JCI, et al. Psychological status in childhood cancer survivors: a report from the Childhood Cancer Survivor Study. J Clin Oncol 2009 May 10;27(14):2396-2404 [FREE Full text] [doi: 10.1200/JCO.2008.21.1433] [Medline: 19255309]
8. Hudson MM, Oeffinger KC, Jones K, Brinkman TM, Krull KR, Mulrooney DA, et al. Age-dependent changes in health status in the Childhood Cancer Survivor cohort. J Clin Oncol 2015 Feb 10;33(5):479-491 [FREE Full text] [doi: 10.1200/JCO.2014.57.4863] [Medline: 25547510]
9. Sender L, Zabokrtsky KB. Adolescent and young adult patients with cancer: a milieu of unique features. Nat Rev Clin Oncol 2015 Aug;12(8):465-480. [doi: 10.1038/nrclinonc.2015.92] [Medline: 26011488]
10. Massey CN, Feig EH, Duque-Serrano L, Wexler D, Moskowitz JT, Huffman JC. Well-being interventions for individuals with diabetes: A systematic review. Diabetes Res Clin Pract 2019 Jan;147:118-133. [doi: 10.1016/j.diabres.2018.11.014] [Medline: 30500545]
11. Folkman S, Moskowitz JT. Positive affect and the other side of coping. Am Psychol 2000 Jun;55(6):647-654. [Medline: 10892207]
12. Chida Y, Steptoe A. Positive psychological well-being and mortality: a quantitative review of prospective observational studies. Psychosom Med 2008 Sep;70(7):741-756. [doi: 10.1097/PSY.0b013e31818105ba] [Medline: 18725425]
13. Moskowitz JT. Positive affect predicts lower risk of AIDS mortality. Psychosom Med 2003;65(4):620-626. [Medline: 12883113]
14. Moskowitz JT, Epel ES, Acree M. Positive affect uniquely predicts lower risk of mortality in people with diabetes. Health Psychol 2008 Jan;27(1S):S73-S82. [doi: 10.1037/0278-6133.27.1.S73] [Medline: 18248108]
15. Pressman SD, Cohen S. Does positive affect influence health? Psychol Bull 2005 Nov;131(6):925-971. [doi: 10.1037/0033-2909.131.6.925] [Medline: 16351329]
16. Huffman JC, DuBois CM, Millstein RA, Celano CM, Wexler D. Positive Psychological Interventions for Patients with Type 2 Diabetes: Rationale, Theoretical Model, and Intervention Development. J Diabetes Res 2015;2015:428349 [FREE Full text] [doi: 10.1155/2015/428349] [Medline: 26064980]
17. Cohn MA, Pietrucha ME, Saslow LR, Hult JR, Moskowitz JT. An online positive affect skills intervention reduces depression in adults with type 2 diabetes. J Posit Psychol 2014 Jan 1;9(6):523-534 [FREE Full text] [doi: 10.1080/17439760.2014.920410] [Medline: 25214877]

18. Peterson JC, Charlson ME, Hoffman Z, Wells MT, Wong S, Hollenberg JP, et al. A randomized controlled trial of positive-affect induction to promote physical activity after percutaneous coronary intervention. Arch Intern Med 2012 Feb 27;172(4):329-336 [FREE Full text] [doi: 10.1001/archinternmed.2011.1311] [Medline: 22269589]
19. Huffman JC, Mastromauro CA, Boehm JK, Seabrook R, Fricchione GL, Denninger JW, et al. Development of a positive psychology intervention for patients with acute cardiovascular disease. Heart Int 2011 Sep 29;6(2):e14 [FREE Full text] [doi: 10.4081/hi.2011.e14] [Medline: 23825741]
20. Boutin-Foster C, Offidani E, Kanna B, Ogedegbe G, Ravenell J, Scott E, et al. Results from the Trial Using Motivational Interviewing, Positive Affect, and Self-Affirmation in African Americans with Hypertension (TRIUMPH). Ethn Dis 2016 Dec 21;26(1):51-60 [FREE Full text] [doi: 10.18865/ed.26.1.51] [Medline: 26843796]
21. Ogedegbe GO, Boutin-Foster C, Wells MT, Allegrante JP, Isen AM, Jobe JB, et al. A randomized controlled trial of positive-affect intervention and medication adherence in hypertensive African Americans. Arch Intern Med 2012 Feb 27;172(4):322-326 [FREE Full text] [doi: 10.1001/archinternmed.2011.1307] [Medline: 22269592]
22. Seligman MEP, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005 Aug;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
23. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
24. Moskowitz JT, Carrico AW, Duncan LG, Cohn MA, Cheung EO, Batchelder A, et al. Randomized controlled trial of a positive affect intervention for people newly diagnosed with HIV. J Consult Clin Psychol 2017 Dec;85(5):409-423 [FREE Full text] [doi: 10.1037/ccp0000188] [Medline: 28333512]
25. Molina Y, Yi JC, Martinez-Gutierrez J, Reding KW, Yi-Frazier JP, Rosenberg AR. Resilience among patients across the cancer continuum: diverse perspectives. Clin J Oncol Nurs 2014 Feb;18(1):93-101 [FREE Full text] [doi: 10.1188/14.CJON.93-101] [Medline: 24476731]
26. Miller KA, Wojcik KY, Ramirez CN, Ritt-Olson A, Freyer DR, Hamilton AS, et al. Supporting long-term follow-up of young adult survivors of childhood cancer: Correlates of healthcare self-efficacy. Pediatr Blood Cancer 2017 Dec;64(2):358-363 [FREE Full text] [doi: 10.1002/pbc.26209] [Medline: 27567026]
27. Rosenberg AR, Yi-Frazier JP, Wharton C, Gordon K, Jones B. Contributors and Inhibitors of Resilience Among Adolescents and Young Adults with Cancer. J Adolesc Young Adult Oncol 2014 Dec 01;3(4):185-193 [FREE Full text] [doi: 10.1089/jayao.2014.0033] [Medline: 25969794]
28. Rosenberg AR, Yi-Frazier JP, Eaton L, Wharton C, Cochrane K, Pihoker C, et al. Promoting Resilience in Stress Management: A Pilot Study of a Novel Resilience-Promoting Intervention for Adolescents and Young Adults With Serious Illness. J Pediatr Psychol 2015 Oct;40(9):992-999 [FREE Full text] [doi: 10.1093/jpepsy/jsv004] [Medline: 25678533]
29. Rosenberg AR, Bradford MC, McCauley E, Curtis JR, Wolfe J, Baker KS, et al. Promoting resilience in adolescents and young adults with cancer: Results from the PRISM randomized controlled trial. Cancer 2018 Oct 01;124(19):3909-3917. [doi: 10.1002/cncr.31666] [Medline: 30230531]
30. Rosenberg AR, Bradford MC, Barton KS, Etsekson N, McCauley E, Curtis JR, et al. Hope and benefit finding: Results from the PRISM randomized controlled trial. Pediatr Blood Cancer 2019 Jan;66(1):e27485. [doi: 10.1002/pbc.27485] [Medline: 30270489]
31. Lauer AL. Treatment of Anxiety and Depression in Adolescents and Young Adults With Cancer. J Pediatr Oncol Nurs 2015 Jan 02;32(5):278-283. [doi: 10.1177/1043454214563406]
32. Skjuve MB, Brandtzæg PB. Chatbots as a new user interface for providing health information to young people. 2018. Social Health Bots URL: https://www.researchgate.net/publication/332859163_Chatbots_as_a_new_user_interface_for_providing_health_information_to_young_people
33. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
34. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
35. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
36. Arnett JJ, Žukauskienė R, Sugimura K. The new life stage of emerging adulthood at ages 18–29 years: implications for mental health. The Lancet Psychiatry 2014 Dec;1(7):569-576. [doi: 10.1016/s2215-0366(14)00080-7]
37. Ethical and Independent Review Services. Independence, MO: Ethical and Independent Review Services; 2010. URL: http://www.eandireview.com/ [accessed 2019-05-03]
38. Moskowitz JT, Hult JR, Duncan LG, Cohn MA, Maurer S, Bussolari C, et al. A positive affect intervention for people experiencing health-related stress: development and non-randomized pilot test. J Health Psychol 2012 Jul;17(5):676-692 [FREE Full text] [doi: 10.1177/1359105311425275] [Medline: 22021272]

39. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
40. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Arlington, VA: Author; 2013:978.
41. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, et al. Clinical validity of PROMIS Depression, Anxiety, and Anger across diverse clinical samples. J Clin Epidemiol 2016 Dec;73:119-127 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.036] [Medline: 26931289]
42. Lazor T, Tigelaar L, Pole JD, De Souza C, Tomlinson D, Sung L. Instruments to measure anxiety in children, adolescents, and young adults with cancer: a systematic review. Support Care Cancer 2017 Dec;25(9):2921-2931. [doi: 10.1007/s00520-017-3743-3] [Medline: 28578534]
43. Clover K, Lambert SD, Oldmeadow C, Britton B, King MT, Mitchell AJ, et al. PROMIS depression measures perform similarly to legacy measures relative to a structured diagnostic interview for depression in cancer patients. Qual Life Res 2018 May;27(5):1357-1367. [doi: 10.1007/s11136-018-1803-x] [Medline: 29423755]
44. HealthMeasures. URL: http://www.healthmeasures.net [accessed 2019-10-01]
45. Fredrickson BL, Cohn MA, Coffey KA, Pek J, Finkel SM. Open hearts build lives: positive emotions, induced through loving-kindness meditation, build consequential personal resources. J Pers Soc Psychol 2008 Nov;95(5):1045-1062 [FREE Full text] [doi: 10.1037/a0013262] [Medline: 18954193]
46. Cohen S, Pressman SD. Positive Affect and Health. Curr Dir Psychol Sci 2016 Jun 22;15(3):122-125. [doi: 10.1111/j.0963-7214.2006.00420.x]
47. Kok BE, Coffey KA, Cohn MA, Catalino LI, Vacharkulksemsuk T, Algoe SB, et al. How Positive Emotions Build Physical Health. Psychol Sci 2013 May 06;24(7):1123-1132. [doi: 10.1177/0956797612470827]
48. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd Ed. Abingdon, UK: Routledge; 2013.
49. Sheard T, Maguire P. The effect of psychological interventions on anxiety and depression in cancer patients: results of two meta-analyses. Br J Cancer 1999 Aug;80(11):1770-1780 [FREE Full text] [doi: 10.1038/sj.bjc.6690596] [Medline: 10468295]
50. Cheung EO, Cohn MA, Dunn LB, Melisko ME, Morgan S, Penedo FJ, et al. A randomized pilot trial of a positive affect skill intervention (lessons in linking affect and coping) for women with metastatic breast cancer. Psychooncology 2016 Nov 16. [doi: 10.1002/pon.4312] [Medline: 27862646]
51. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 13;5(4):e64. [doi: 10.2196/mental.9782]
52. Torre JB, Lieberman MD. Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. Emotion Review 2018 Mar 20;10(2):116-124. [doi: 10.1177/1754073917742706]
53. Schueller SM, Munoz RF, Mohr DC. Realizing the Potential of Behavioral Intervention Technologies. Current Directions in Psychological Science 2013 Dec 03;22(6):478-483. [doi: 10.1177/0963721413495872]
54. Wozney L, Huguet A, Bennett K, Radomski AD, Hartling L, Dyson M, et al. How do eHealth Programs for Adolescents With Depression Work? A Realist Review of Persuasive System Design Components in Internet-Based Psychological Therapies. J Med Internet Res 2017 Aug 09;19(8):e266 [FREE Full text] [doi: 10.2196/jmir.7573] [Medline: 28793983]
55. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental Healthcare Technologies for Anxiety and Depression in Childhood and Adolescence: Systematic Review of Studies Reporting Implementation Outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
56. Ramo DE, Thrul J, Delucchi KL, Hall S, Ling PM, Belohlavek A, et al. A randomized controlled evaluation of the tobacco status project, a Facebook intervention for young adults. Addiction 2018 May 24. [doi: 10.1111/add.14245] [Medline: 29797621]
57. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study. Pediatr Blood Cancer 2017 Dec;64(12). [doi: 10.1002/pbc.26660] [Medline: 28618158]
58. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018 Internet.: Pew Research Center; 2019 Apr 10. Factank: News in the Numbers URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-06-03]
59. Hernandez R, Burrows B, Wilund K, Cohn M, Xu S, Moskowitz JT. Feasibility of an Internet-based positive psychological intervention for hemodialysis patients with symptoms of depression. Soc Work Health Care 2018;57(10):864-879. [doi: 10.1080/00981389.2018.1523268] [Medline: 30277449]

Abbreviations

PRISM: Promoting Resilience in Stress Management
PROMIS: Patient-Reported Outcomes Measurement Information System

UCLA: University of California, Los Angeles

Edited by G Eysenbach; submitted 12.06.19; peer-reviewed by P Klainin-Yobas, S Gow Jen; comments to author 04.07.19; revised version received 05.09.19; accepted 24.09.19; published 31.10.19.

Please cite as:
Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J
Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial
JMIR Mhealth Uhealth 2019;7(10):e15018
URL: http://mhealth.jmir.org/2019/10/e15018/
doi:10.2196/15018
PMID:31674920

Original Paper

Dropout and Abstinence Outcomes in a National Text Messaging Smoking Cessation Intervention for Pregnant Women, SmokefreeMOM: Observational Study

Kristyn Kamke[1], MS, PhD; Emily Grenen[2], MSc; Cendrine Robinson, MPH, PhD; Sherine El-Toukhy[1], MA, PhD

[1]Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States

[2]ICF, Rockville, MD, United States

Corresponding Author:
Sherine El-Toukhy, MA, PhD
Division of Intramural Research
National Institute on Minority Health and Health Disparities
National Institutes of Health
7201 Wisconsin Ave
RM 533M
Bethesda, MD, 20814
United States
Phone: 1 301 594 4743
Email: sherine.el-toukhy@nih.gov

Abstract

Background: Population-level text messaging smoking cessation interventions may reduce racial and ethnic differences in smoking among pregnant women.

Objective: Our objective was to examine racial and ethnic differences in dropout, response, and abstinence rates among users of a US national, publicly available text messaging cessation intervention targeting pregnant women, SmokefreeMOM.

Methods: Participants were online subscribers to SmokefreeMOM who set a prospective quit date within the 9 months before their due date. We examined demographics, smoking frequency, number of cigarettes smoked per day, and prequit time (up to 14 days of preparation time before quit date) as correlates of response rate and abstinence at 8 time points: quit date, day 7, day 14, day 21, day 28, day 35, day 42 (intervention end), and day 72 (1-month follow-up). We conducted survival analysis of time from quit date to dropout by race and ethnicity.

Results: The mean age of the analytic sample of 1288 users was 29.46 (SD 7.11) years. Of these, 65.81% (848/1288) were white, 16.04% (207/1288) were black, 8.86% (114/1288) were Latina, and 9.29% (120/1288) were multiracial, American Indian/Alaska Native, Native Hawaiian Pacific Islander, or other; 82.68% (1065/1288) had some college education or less. Point-prevalence abstinence was 14.51% (157/1082) on quit day, 3.51% (38/1082) at intervention end, and 1.99% (21/1053) at 1-month follow-up. Black users (hazard ratio 0.68, 95% CI 0.51-0.91) and those with a high school degree or less (hazard ratio 0.66, 95% CI 0.49-0.89) or some college education (hazard ratio 0.75, 95% CI 0.57-0.99) were less likely to drop out than whites or users with a bachelor's degree or higher. Response and abstinence rates were similar across race, ethnicity, and education.

Conclusions: Enrollment was low among racial and ethnic minority women but high among less-educated women. Abstinence at intervention end and 1-month follow-up was lower than that in controlled trials of text messaging cessation interventions for pregnant women (range 7%-20%). Increasing the reach, engagement, and effectiveness of SmokefreeMOM, especially among women with high rates of smoking during pregnancy, must be prioritized.

(JMIR Mhealth Uhealth 2019;7(10):e14699) doi:10.2196/14699

KEYWORDS

smoking cessation; pregnancy; women's health; mHealth; text messaging intervention

Introduction

Background

Pregnant women are a priority population for smoking cessation efforts [1,2]. Smoking during pregnancy confers negative health outcomes on the mother, fetus, and infant, including preterm birth, low birthweight, birth defects, and infant death [1,3]. Nevertheless, 6.9% (264,920) of pregnant women in the United States reported smoking in 2017, of whom 79% continued to smoke throughout pregnancy [4]. Further, evidence suggests that 31% to 52% of women will resume smoking postnatally [5], exposing their babies to secondhand smoke-related problems [6].

Racial and ethnic disparities in smoking during pregnancy are apparent. American Indian/Alaska Native (AI/AN) (16.4%) and white (10.1%) women have the highest smoking rates, exceeding the US national average of smoking during pregnancy [4]. Black and Native Hawaiian and Pacific Islander (NHPI) women have the next highest rates (5.6% and 4.6%, respectively), followed by Latina and Asian women (1.8% and 0.5%, respectively) [4]. Importantly, pregnant women, especially racial and ethnic minorities, may underreport their smoking status. Biological measures show that nicotine exposure among black and Latina pregnant women is 2 to 4 times higher than self-reported rates [7]. Whereas rates of smoking during pregnancy in the United States decreased 4% to 5% annually among white and Latina women from 1985 to 2013, rates have not decreased among black women over time [8]. Additionally, smoking cessation rates increased 9% annually from 1985 to 2013 among Latina women but not among white or black women [8].

Other risk factors can exacerbate the adverse health outcomes of smoking during pregnancy among minorities. Greater psychosocial stress and experiences of discrimination among racial and ethnic minority women heighten the risk of smoking during pregnancy and negatively affect prenatal and postnatal health [9,10]. Indeed, compared with white women, black, Latina, AI/AN, and NHPI women experience higher rates of smoking-related negative birth outcomes, including preterm birth, lower birthweight, and infant death [4,11]. Compared with white women, black, Latina, AI/AN, and NHPI women are less likely to begin prenatal care in the first trimester [4,11], limiting time for health practitioners to intervene in their smoking behavior, and are less likely to breastfeed [4,12], a key protective factor against smoking during and after pregnancy [13].

Behavioral interventions are necessary cessation aids for pregnant women, as the effectiveness and safety of nicotine replacement therapy during pregnancy is not well established [1,14]. Population-level short message service (SMS) text messaging interventions have been used widely and effectively for smoking cessation [15,16]. They can be tailored to deliver pregnancy-focused information and have high penetration, allowing vulnerable groups to access cessation services typically unavailable to them due to financial or logistical constraints [17,18]. These interventions also reduce pregnancy-specific barriers to seeking smoking cessation resources, such as fear of stigmatization or legal repercussions [17,18].

SMS text messaging smoking cessation interventions are efficacious among the general population of smokers [15,16]. However, few SMS text messaging interventions have been developed for and evaluated among pregnant women [17]. One available intervention is SmokefreeMOM, an evidence-informed SMS text messaging smoking cessation intervention for pregnant women. SmokefreeMOM is offered within a suite of Web- and mobile-based smoking cessation resources, accessible through Smokefree.gov, a US National Cancer Institute initiative [19].

Objective

Evidence shows that SmokefreeMOM is acceptable and engaging, improves motivation for quitting, reduces craving symptoms, and promotes smoking abstinence [20,21]. However, to our knowledge, no study has evaluated the population-level implementation of SmokefreeMOM. Of interest is whether SmokefreeMOM yields comparable retention, response, and abstinence rates among racial and ethnic subgroups of pregnant smokers. Thus, this study examined racial and ethnic differences in dropout, response, and abstinence rates among SmokefreeMOM users.

Methods

Intervention Description

SmokefreeMOM is a free, publicly available smoking cessation intervention, accessible to anyone in the United States with an SMS text messaging–enabled mobile phone. It was developed with input from pregnant smokers and is grounded in social cognitive theory and other proven behavioral strategies for smoking cessation [22-24]. Women sign up for SmokefreeMOM online at Smokefree.gov or by texting the keyword “Mom” to 222888. Upon enrollment, users select their goal for the program: “I want to quit smoking” or “I am not ready to make a change but would like to receive messages on smoking and health.” A previous iteration also allowed users to select “I want to cut back.” Users who are not ready to quit or want to cut back receive an alternative message library. Users who want to quit smoking set a quit date recommended to be within 14 days of sign-up and answer demographic and smoking-related questions. Women can reset their quit date and restart the program by texting “DATE” at any time. Prior to the quit date (ie, prequit period), users receive up to 34 messages over a maximum of 14 days to prepare them to quit smoking. The self-set quit date triggers the 42-day SmokefreeMOM intervention with 101 smoking cessation messages, including behavioral challenges, facts on the effects of smoking on a baby’s development, advice from former pregnant smokers, and links to smoking cessation resources. Users receive up to 313 days of messages related to maternal and child health that correspond with their due date, continuing beyond the 42-day intervention.

Study Population

In 2764 user records available from April 2014 to June 2018, 584 users were not asked the race, ethnicity, and education questions because they enrolled via mobile keyword. This reduced our initial sample to 2180. We devised our exclusion criteria around SmokefreeMOM parameters (Table 1). First, for users who reset their quit date, we used data from their most

recent quit date and excluded prior records for each user. Second, we excluded users who were not ready or did not want to quit smoking, had no quit date or due date, retrospectively set a quit date, reported a due date before their quit date, or reported a due date that was more than 280 days after their sign-up date (ie, time from sign-up to due date), which is the maximum gestational period [25]. We also excluded users who dropped out of the study prior to the prequit period without receiving any intervention dose. Finally, we downloaded data on June 30, 2018, which we deemed to be the end-of-study date. Accordingly, we excluded users whose quit date was less than 42 days prior to the end-of-study date because they did not have the opportunity to complete the intervention. The final analytic sample comprised 1288 users.

Table 1. Exclusion criteria (N=2180 Web-enrolled users in SmokefreeMOM between April 2014 and June 2018).

First reason for exclusion	N (%)[a]
Multiple sign-ups	458 (21.01)
Not ready to quit	146 (6.70)
No quit date	28 (1.28)
Quit date precedes sign-up date	60 (2.75)
Due date is more than 280 days after sign-up date	32 (1.47)
No due date	1 (0.05)
Due date precedes quit date	101 (4.63)
Opted out before prequit period	12 (0.55)
Quit date less than 42 days before end of study[b]	54 (2.48)
Total excluded from study	892 (40.92)
Total included in study	1288 (59.08)

[a]N (%) for multiple sign-ups refers to records rather than unique participants; number of records per user ranged from 2 to 37 (mean 2.92, SD 3.08); n (%) for all other exclusion criteria represents unique users excluded.

[b]End-of-study date was June 30, 2018, when the data were pulled.

Measures

At sign-up, users reported their age, race, ethnicity, educational attainment, smoking frequency, cigarettes smoked per day, zip code, and whether they used a Web-enabled phone. We combined race and ethnicity, in which we classified ethnically Latina women as such regardless of their race and classified all others by their race (eg, black). Users reported their due date and selected a quit date (if their goal was to quit smoking). Three dates were captured automatically: (1) sign-up, (2) opt-out if user texted "STOP" or if text messages were undeliverable to the registered phone number, and (3) quit date reset, if any. We used self-reported and automatically captured dates to derive the following time variables: (1) time from sign-up to quit date, (2) prequit time, and (3) time from quit date to dropout.

We used time from sign-up to quit date to derive a continuous prequit-time variable, which we used as a covariate in our analyses because of associations between preparation stage and intervention dose with smoking cessation [16,26]. Although users were prompted to set a quit date within 14 days of sign-up, some users selected a quit date beyond the maximum recommended prequit period, during which they received no intervention messages. For these users, we reset their prequit time to 14 days. We made no changes for those who did not exceed 14 days. For example, if a woman set a quit date 30 days after sign-up, we reset her prequit time to 14 days; however, if a woman set her quit date at 7 days after sign-up, her prequit time remained at 7 days. Time from quit date to dropout ranged from –14 (for women who dropped out 14 days before their quit date on the first day that prequit intervention content was delivered) to 42 (for women who dropped out on the last day of the intervention).

We derived a binary dropout variable from the time from quit date to dropout variable, in which we considered users to be *noncompleters* if they texted "STOP" to opt out or for whom text messages were undeliverable by the intervention between day –14 and intervention end (day 42), or to be *completers* if they were retained beyond the intervention end, regardless of their engagement with the intervention. SmokefreeMOM prompted users via text message to report their smoking status every week starting on the quit date and ending on intervention end day. They also reported their smoking status at 1, 3, and 6 months after intervention end (ie, days 72, 132, and 222). Smoking status was captured by a yes/no question: "Hi there! Have you smoked in the last 7 days?" We derived a binary responding status variable, in which we deemed users to be *responders* if they responded to the smoking status prompt with yes or no and as *nonresponders* if they did not respond.

Data Analysis

Almost one-third of users (382/1288, 29.66%) had missing values on 1 or more user characteristics, suggesting listwise deletion would bias results [27]. The Little chi-square test [28] showed that data were not missing completely at random (χ^2_{81}=117.8, *P*=.01). Statistical analyses are likely to be biased when more than 10% of data are missing [29,30]. Since multiple imputation methods are preferred when data are not missing completely at random [30], we imputed missing data (n=20)

for race, ethnicity, educational attainment, smoking frequency, cigarettes per day, and region with input data that met our inclusion criteria. Because we imputed categorical variables, we specified a logistic prediction model and used a generalized logit model for nonordinal variables (ie, race, region). Although we have a combined race and ethnicity variable, we imputed race and ethnicity separately because some users were missing only 1 of the 2 variables. Age, ownership of a Web-enabled phone, and prequit time had no missing data points and were used as covariates in multiple imputations models. We winsorized self-reported age that fell outside 10 to 54 years (17/1288, range 55-99) [31] to preserve data from outliers while minimizing their impact on the results [32]. This age range is consistent with the US Centers for Disease Control and Prevention's range of probable ages for becoming pregnant [31].

Using imputed data, we conducted a Cox regression survival analysis of time from quit date to dropout by race and ethnicity. We did not detect violations of multicollinearity or proportionality of hazards assumptions. Users were right-censored at 42 days after quit date. We conduced logistic regressions to examine correlates of response and abstinence rates on quit date through 1-month follow-up. We limited response rate models to users who remained in the intervention and had the opportunity to respond to each smoking status prompt; thus, the n for response rate models varied at each assessment time. We based abstinence models on an intent-to-treat approach in which we considered users who did not respond to the smoking status prompt or had dropped out before the date of prompt to be smokers. Cox regression survival analysis and abstinence models included users who did not drop out prior to the quit date and had the opportunity to respond to the first smoking status prompt on the quit date (n=1082). We adjusted dropout, response rate, and abstinence models for age, race and ethnicity, educational attainment, region, smoking frequency, cigarettes smoked per day, and prequit time.

Results

Participant Characteristics

Among the analytic sample, 65.81% (848/1288) were white, 16.04% (207/1288) were black, 8.86% (114/1288) were Latina, and 9.29% (120/1288) were multiracial, Asian, AI/AN, NHPI, or other race (Tables 2 and 3; Multimedia Appendix 1 shows complete case user characteristics). Approximately 17.32% (223/1288) had a bachelor's degree or higher, whereas 82.68% (1065/1288) had some college education, or a high school degree or less. On average, users were 29.46 years old and signed up for SmokefreeMOM at the beginning of their second trimester (ie, 3.6 months pregnant) and had around 5.4 months until their due date.

Dropout Rates

Of all SmokefreeMOM users, 15.99% (206/1288) dropped out before their quit date and 39.52% (509/1288) dropped out on or after their quit date. Compared with white users who remained in the program until their quit date, black users were less likely to drop out of SmokefreeMOM before intervention end (hazard ratio [HR] 0.68, 95% CI 0.51-0.91, Figure 1, Table 4). At the mean of the covariates, the survival rate was around 80% for black women on day 7 and around 62% on day 42 (Figure 1). Women with some college education (HR 0.75, 95% CI 0.57-0.99) and those with high school education or less (HR 0.66, 95% CI 0.49-0.89) were less likely to drop out of SmokefreeMOM than were women with bachelor's degrees or higher. A longer prequit time (ie, 0-14 days of preparation time before quit date) was associated with a lower likelihood of dropping out (HR 0.97, 95% CI 0.96-0.99). A logistic regression analysis for dropout that included women who dropped out before and on or after the quit day showed similar results (Multimedia Appendix 2).

Table 2. SmokefreeMOM user characteristics, imputed data (N=1288).

Characteristics	Total	Noncompleters[a]	Completers[b]
Users, n (%)	1288 (100.00)[c]	715 (55.51)[c]	573 (44.49)[c]
Age, years			
Mean (SD)	29.58 (7.64)	29.64 (7.68)	29.49 (7.60)
5% trimmed mean (SD)	29.02 (6.35)[d]	29.09 (6.33)[e]	28.93 (6.38)[f]
Range	14-99	16-99	14-72
Median (IQR[g])	29.00 (10.00)	29.00 (10.00)	28.00 (10.00)
Age (winsorized), years			
Mean (SD)	29.46 (7.11)	29.52 (7.05)	29.39 (7.19)
5% trimmed mean (SD)	29.02 (6.35)[d]	29.09 (6.33)[e]	28.93 (6.38)[f]
Range	14-54	16-54	14-54
Median (IQR)	29.00 (10.00)	29.00 (10.00)	28.00 (10.00)
Time from sign-up date to due date, days			
Mean (SD)	161.21 (73.65)	164.73 (75.57)	156.82 (71.01)
5% trimmed mean (SD)	164.49 (75.93)[d]	168.45 (78.25)[e]	159.47 (72.34)[f]
Range	0-279	0-279	0-279
Median (IQR)	175.50 (117.00)	184.00 (113.00)	167.00 (113.00)
Time from sign-up to quit date, days			
Mean (SD)	10.41 (15.74)	9.38 (14.06)	11.70 (17.53)
5% trimmed mean (SD)	8.04 (9.60)[d]	7.58 (7.65)[e]	8.93 (12.65)[f]
Range	0-197	0-197	0-144
Median (IQR)	7.00 (13.00)	7.00 (13.00)	7.00 (13.00)
Prequit time, days			
Mean (SD)	6.12 (5.53)	5.20 (5.29)	7.28 (5.61)
5% trimmed mean (SD)	6.03 (5.83)[d]	5.00 (5.58)[e]	7.31 (5.91)[f]
Range	0-14	0-14	0-14
Median (IQR)	5.00 (12.00)	3.00 (10.00)	7.00 (13.00)
Time from quit day to dropout, days[h]			
Mean (SD)	10.70 (11.74)[i]	10.70 (11.74)[i]	N/A[j]
5% trimmed mean (SD)	9.67 (11.92)[k]	9.67 (11.92)[k]	N/A
Range	0-42[i]	0-42[i]	N/A
Median (IQR)	6.00 (16.00)[i]	6.00 (16.00)[i]	N/A

[a]Noncompleters were users who opted out of the intervention any time on or between day –14 (14 days before quit date) and intervention end (day 42).
[b]Completers were users who remained in SmokefreeMOM until after intervention end.
[c]Imputed Ns have 20 records and thus n is 1/20th of a subject rounded to the nearest integer.
[d]N=1158.
[e]n=643.
[f]n=515.
[g]IQR: interquartile range.
[h]Among those who made it to their quit date.
[i]N=509.
[j]Not applicable.

[k]N=457.

Table 3. SmokefreeMOM user characteristics, imputed data (N=1288).

Characteristics	Total, n (%)[a]	Noncompleters[b], n (%)[a]	Completers[c], n (%)[a]
Race and ethnicity			
White	848 (65.81)	499 (69.80)	349 (60.84)
Black	207 (16.04)	88 (12.24)	119 (20.78)
Latina	114 (8.86)	69 (9.62)	45 (7.91)
Multiracial, Asian, AI/AN[d], NHPI[e], other[f]	120 (9.29)	60 (8.34)	60 (10.48)
Education			
High school or less	513 (39.86)	267 (37.32)	247 (43.02)
Some college	552 (42.82)	302 (42.17)	250 (43.63)
College graduate or higher	223 (17.32)	147 (20.50)	77 (13.35)
Region[g]			
Northeast	180 (13.94)	105 (14.69)	75 (13.00)
Midwest	332 (25.74)	201 (28.06)	131 (22.84)
South	569 (44.15)	293 (41.01)	275 (48.07)
West	208 (16.17)	116 (16.24)	92 (16.08)
Smoking frequency			
Nondaily	143 (11.09)	67 (9.39)	76 (13.21)
Daily	1145 (88.91)	648 (90.61)	497 (86.79)
Cigarettes per day			
Light (<10 cigarettes)	762 (59.12)	397 (55.57)	364 (63.55)
Moderate (11-19 cigarettes)	416 (32.29)	240 (33.60)	176 (30.65)
Heavy (≥20 cigarettes)	111 (8.59)	77 (10.83)	33 (5.79)
Web-enabled phone			
Yes	1238 (96.12)	687 (96.08)	551 (96.16)
No	50 (3.88)	28 (3.92)	22 (3.84)
Time of dropout[h]			
Prior to quit date	206 (15.99)	206 (15.99)	N/A[i]
On or after quit date	509 (39.52)	509 (39.52)	N/A

[a]Imputed Ns have 20 records and thus n is 1/20th of a subject rounded to the nearest integer.

[b]Noncompleters were users who opted out of the intervention any time on or between day –14 (14 days before quit date) and intervention end (day 42).

[c]Completers were users who remained in SmokefreeMOM until after intervention end.

[d]AI/AN: American Indian/Alaska Native.

[e]NHPI: Native Hawaiian and Pacific Islander.

[f]Complete case sample was 4.72% (58/1230) multiracial, 1.14% (14/1230) Asian, 0.73% (9/1230) AI/AN, 0.49% (6/1230) NHPI, and 2.11% (26/1230) other.

[g]Users provided their zip codes, which were automatically converted into US state. We categorized states into US Census Bureau region. We categorized 1 user who lived in Puerto Rico, for which there is no census region, into South to retain her data in the analyses.

[h]The remaining 44.49% (573/1288) did not drop out prior to intervention end.

[i]Not applicable.

Figure 1. Survival analysis of days in SmokefreeMOM by race and ethnicity adjusted for age (winsorized), educational attainment, region, smoking frequency, cigarettes per day, and prequit time, imputed data (N=1082). AI/AN: American Indian/Alaska Native; NHPI: Native Hawaiian and Pacific Islander.

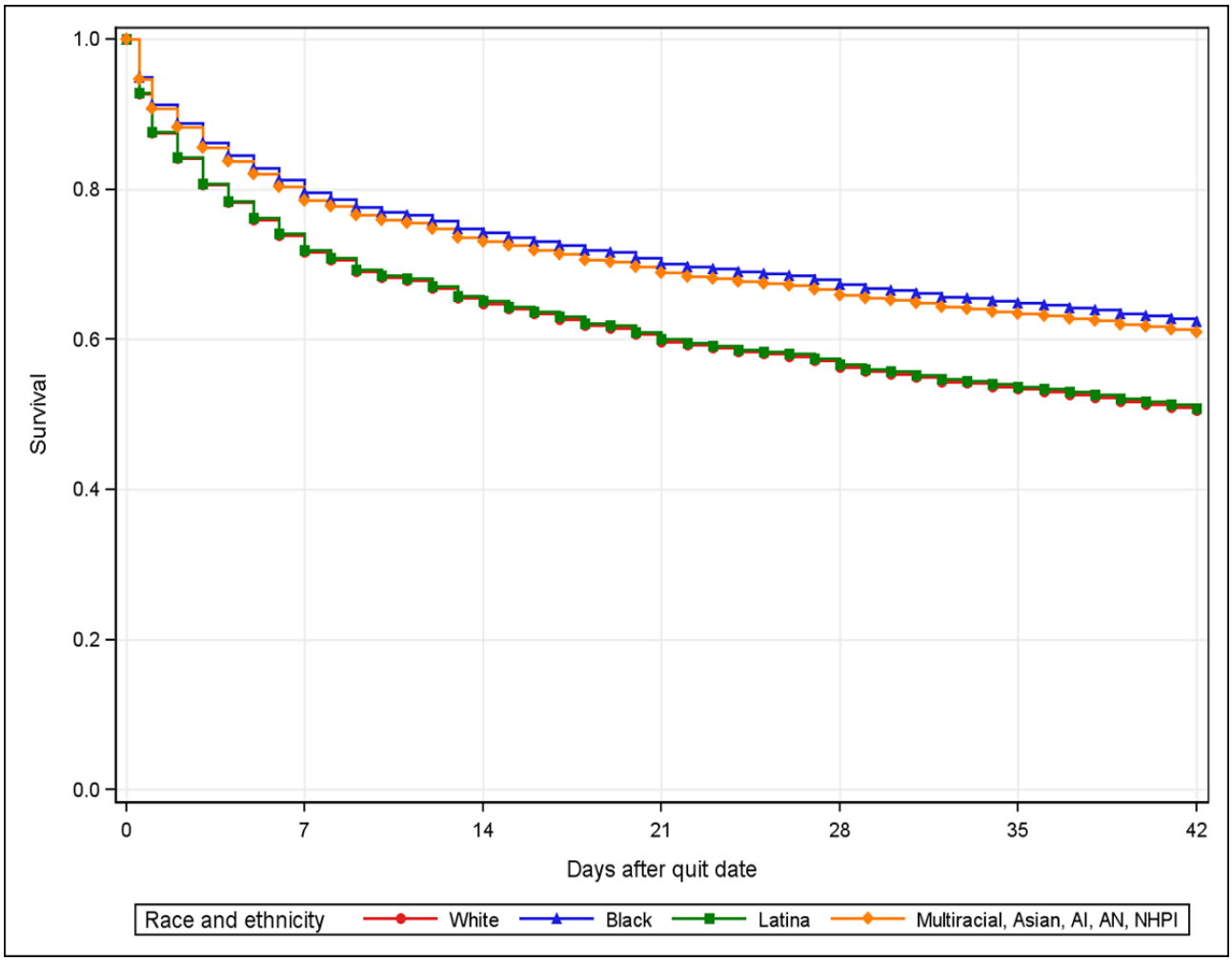


Table 4. Hazard ratios of dropping out by user characteristics, imputed data (N=1082).

Characteristics	Hazard ratio	95% CI	*P* value
Age (winsorized)	0.99	0.97-1.00	.11
Race and ethnicity (reference: white)			
Black	0.68	0.51-0.91	.009
Latina	0.99	0.72-1.37	.95
Multiracial, Asian, AI/AN[a], NHPI[b], other	0.72	0.51-1.01	.06
Education (reference: college graduate or higher)			
High school or less	0.66	0.49-0.89	.006
Some college	0.75	0.57-0.99	.04
Region (reference: South)			
Northeast	1.15	0.87-1.51	.32
Midwest	1.20	0.96-1.49	.10
West	1.11	0.85-1.44	.44
Cigarettes per day (reference: light)			
Moderate	1.03	0.85-1.26	.74
Heavy	1.38	0.99-1.92	.06
Smoking frequency (reference: nondaily)			
Daily	1.32	0.96-1.80	.08
Prequit time	0.97	0.96-0.99	.001

[a]AI/AN: American Indian/Alaska Native.

[b]NHPI: Native Hawaiian and Pacific Islander.

Response and Abstinence Rates

Response rates averaged 29.48% (319/1082) on quit date, were between 10.23% (62/606) and 18.45% (150/813) through day 35, then dropped below 9.86% (57/578) for intervention end and 1-month follow-up. Overall abstinence ranged from 14.51% (157/1082) on quit day to 3.51% (38/1082) at intervention end and 1.99% (21/1053) at 1-month follow-up (Multimedia Appendix 3). Although no clear pattern of associations emerged between user characteristics and either response rate or abstinence (Tables 5-8), some significant associations are noteworthy. Black users (adjusted odds ratio [aOR] 0.36, 95% CI 0.15-0.91), those with high school education or less (aOR 0.29, 95% CI 0.15-0.56), and daily smokers (aOR 0.54, 95% CI 0.31-0.94) were less likely to respond at individual time points throughout the intervention than were their respective reference groups. Similarly, users with high school education or less (vs those with bachelor's degrees or higher; aOR 0.39, 95% CI 0.17-0.92) were less likely to be abstinent on day 7, moderate smokers (vs light smokers) were more likely to be abstinent on day 35 (aOR 2.31, 95% CI 1.12-4.75), and daily smokers (vs nondaily) were less likely to be abstinent on quit day (aOR 0.42, 95% CI 0.26-0.67), day 14 (aOR 0.38, 95% CI 0.20-0.74), and 1-month follow-up (aOR 0.27, 95% CI 0.08-0.89). Longer prequit intervention time was associated with a reduced likelihood of abstinence on quit day (aOR 0.96, 95% CI 0.93-0.99). Sensitivity analyses using unimputed data with listwise deletion showed fairly comparable results (Multimedia Appendix 4).

Table 5. Associations between user characteristics and response rates: from quit date to days 7-21, imputed data.

Characteristics	Quit date (n=1082)[a]		Day 7 (n=813)[a]		Day 14 (n=727)[a]		Day 21 (n=678)[a]	
	aOR[b] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Age (winsorized)	0.99 (0.97-1.01)	.27	0.98 (0.95-1.01)	.25	0.99 (0.97-1.03)	.73	1.00 (0.97-1.04)	.80
Race and ethnicity (reference: white)								
Black	1.05 (0.71-1.54)	.82	0.85 (0.49-1.45)	.55	0.87 (0.50-1.53)	.64	1.01 (0.57-1.78)	.97
Latina	0.58 (0.33-1.01)	.05	0.63 (0.28-1.41)	.26	1.22 (0.56-2.62)	.62	0.35 (0.10-1.19)	.09
Multiracial, Asian, AI/AN[c], NHPI[d], other	0.86 (0.53-1.39)	.53	0.86 (0.47-1.60)	.64	1.61 (0.87-2.97)	.13	1.04 (0.49-2.22)	.91
Education (reference: college degree or higher)								
High school or less	0.64 (0.40-1.02)	.06	0.29 (0.15-0.56)	<.001	0.56 (0.26-1.20)	.14	0.73 (0.30-1.75)	.47
Some college	0.98 (0.65-1.48)	.93	0.64 (0.38-1.06)	.08	1.10 (0.57-2.14)	.77	1.45 (0.65-3.22)	.36
Region (reference: South)								
Northeast	1.20 (0.80-1.82)	.38	1.25 (0.71-2.17)	.44	1.21 (0.68-2.16)	.52	1.12 (0.58-2.15)	.74
Midwest	1.13 (0.81-1.58)	.47	0.96 (0.59-1.53)	.85	1.41 (0.88-2.28)	.16	1.44 (0.87-2.38)	.15
West	1.26 (0.85-1.85)	.25	1.32 (0.78-2.23)	.30	1.02 (0.56-1.85)	.94	0.48 (0.21-1.06)	.07
Cigarettes per day (reference: light)								
Moderate	1.21 (0.89-1.65)	.22	1.18 (0.78-1.78)	.45	1.49 (0.96-2.32)	.07	0.91 (0.55-1.50)	.71
Heavy	0.92 (0.52-1.63)	.78	0.64 (0.24-1.74)	.38	1.13 (0.46-2.74)	.79	1.58 (0.65-3.83)	.32
Smoking frequency (reference: nondaily)								
Daily	0.69 (0.46-1.05)	.08	0.71 (0.42-1.19)	.19	0.54 (0.31-0.94)	.03	0.91 (0.47-1.76)	.77
Prequit time	0.99 (0.97-1.02)	.51	0.97 (0.94-1.00)	.08	1.01 (0.98-1.05)	.50	0.99 (0.95-1.03)	.66

[a]Number of users who had not dropped out of the intervention and had the opportunity to respond (or not) each time smoking status was assessed.

[b]aOR: adjusted odds ratio.

[c]AI/AN: American Indian/Alaska Native.

[d]NHPI: Native Hawaiian and Pacific Islander.

Table 6. Associations between user characteristics and response rates: from days 28-72, imputed data.

Characteristics	Day 28 (n=642)[a]		Day 35 (n=606)[a]		Day 42 (n=578)[a]		Day 72 (n=535)[a]	
	aOR[b] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Age (winsorized)	0.99 (0.96-1.03)	.65	1.01 (0.97-1.05)	.65	1.01 (0.97-1.05)	.78	0.98 (0.93-1.04)	.58
Race and ethnicity (reference: white)								
Black	1.15 (0.62-2.16)	.65	0.36 (0.15-0.91)	.03	0.67 (0.30-1.49)	.32	0.86 (0.29-2.52)	.78
Latina	0.63 (0.21-1.89)	.41	0.53 (0.15-1.83)	.31	0.93 (0.30-2.89)	.90	1.03 (0.27-3.90)	.96
Multiracial, Asian, AI/AN[c], NHPI[d], other	0.94 (0.41-2.17)	.89	0.79 (0.31-2.02)	.63	1.34 (0.54-3.32)	.52	1.38 (0.46-4.13)	.56
Education (reference: college degree or higher)								
High school or less	0.55 (0.22-1.40)	.21	0.73 (0.26-2.07)	.56	0.70 (0.26-1.94)	.50	0.59 (0.15-2.38)	.46
Some college	1.09 (0.49-2.44)	.83	1.48 (0.57-3.83)	.42	1.39 (0.54-3.54)	.49	0.92 (0.27-3.19)	.90
Region (reference: South)								
Northeast	1.16 (0.58-2.34)	.68	0.84 (0.36-1.97)	.69	1.19 (0.52-2.74)	.68	0.79 (0.24-2.58)	.70
Midwest	0.66 (0.34-1.26)	.21	0.89 (0.45-1.75)	.73	1.44 (0.75-2.77)	.27	0.81 (0.30-2.16)	.67
West	0.82 (0.39-1.71)	.60	0.85 (0.38-1.91)	.69	0.39 (0.13-1.18)	.10	1.26 (0.47-3.36)	.64
Cigarettes per day (reference: light)								
Moderate	1.09 (0.63-1.90)	.75	1.50 (0.83-2.71)	.18	0.89 (0.46-1.70)	.72	1.90 (0.83-4.33)	.13
Heavy	0.78 (0.22-2.79)	.71	0.60 (0.13-2.75)	.51	0.86 (0.24-3.14)	.82	0.84 (0.10-6.96)	.87
Smoking frequency (reference: nondaily)								
Daily	0.60 (0.31-1.17)	.14	0.62 (0.29-1.34)	.23	0.72 (0.32-1.60)	.42	0.36 (0.15-0.91)	.03
Prequit time	0.99 (0.94-1.03)	.54	1.00 (0.95-1.05)	.92	1.01 (0.96-1.06)	.73	0.95 (0.88-1.01)	.10

[a]Number of users who had not dropped out of the intervention and had the opportunity to respond (or not) each time smoking status was assessed; on day 72, in addition to dropouts on quit day through day 42, we excluded users whose quit days were 43-71 days before the end of the study and those who opted out on days 43-71 because they did not receive the day 72 prompt.

[b]aOR: adjusted odds ratio.

[c]AI/AN: American Indian/Alaska Native.

[d]NHPI: Native Hawaiian and Pacific Islander.

Table 7. Associations between user characteristics and abstinence rates: from quit date to days 7-21, imputed data.

Characteristics	Quit date (n=1082)[a]		Day 7 (n=1082)[a]		Day 14 (n=1082)[a]		Day 21 (n=1082)[a]	
	aOR[b] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Age (winsorized)	1.01 (0.99-1.04)	.34	1.00 (0.96-1.04)	.88	1.01 (0.97-1.05)	.63	1.00 (0.95-1.04)	.95
Race and ethnicity (reference: white)								
Black	1.27 (0.78-2.05)	.34	1.11 (0.55-2.22)	.77	1.16 (0.59-2.29)	.67	1.32 (0.64-2.70)	.45
Latina	0.72 (0.35-1.47)	.37	0.78 (0.29-2.09)	.62	0.98 (0.36-2.62)	.96	0.22 (0.03-1.62)	.14
Multiracial, Asian, AI/AN[c], NHPI[d], other	0.97 (0.52-1.80)	.92	1.26 (0.58-2.74)	.57	1.01 (0.41-2.50)	.99	0.80 (0.27-2.34)	.68
Education (reference: college degree or higher)								
High school or less	0.59 (0.33-1.06)	.08	0.39 (0.17-0.92)	.03	0.57 (0.23-1.38)	.21	0.87 (0.24-3.14)	.84
Some college	0.94 (0.56-1.57)	.80	0.82 (0.40-1.65)	.57	1.05 (0.49-2.27)	.90	1.41 (0.48-4.15)	.53
Region (reference: South)								
Northeast	0.83 (0.48-1.46)	.52	0.76 (0.35-1.63)	.47	0.66 (0.29-1.47)	.31	1.10 (0.50-2.43)	.81
Midwest	0.99 (0.64-1.53)	.96	0.42 (0.20-0.88)	.02	0.75 (0.40-1.41)	.37	0.73 (0.36-1.47)	.38
West	1.31 (0.81-2.14)	.27	1.10 (0.57-2.13)	.77	0.60 (0.27-1.36)	.22	0.42 (0.14-1.24)	.12
Cigarettes per day (reference: light)								
Moderate	1.14 (0.76-1.71)	.53	0.93 (0.52-1.66)	.80	1.17 (0.65-2.11)	.61	1.02 (0.54-1.91)	.96
Heavy	1.06 (0.51-2.21)	.87	0.20 (0.03-1.49)	.12	0.75 (0.22-2.60)	.66	0.52 (0.12-2.27)	.38
Smoking frequency (reference: nondaily)								
Daily	0.42 (0.26-0.67)	<.001	0.60 (0.31-1.18)	.14	0.38 (0.20-0.74)	.004	1.23 (0.46-3.25)	.68
Prequit time	0.96 (0.93-0.99)	.02	0.96 (0.92-1.01)	.08	0.99 (0.95-1.04)	.73	0.98 (0.94-1.03)	.50

[a]N=1082 reflects total users who made it to quit date.

[b]aOR: adjusted odds ratio.

[c]AI/AN: American Indian/Alaska Native.

[d]NHPI: Native Hawaiian and Pacific Islander.

Table 8. Associations between user characteristics and abstinence rates: from days 28-72, imputed data.

Characteristics	Day 28 (n=1082)[a]		Day 35 (n=1082)[a]		Day 42 (n=1082)[a]		Day 72 (n=1053)[a]	
	aOR[b] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Age (winsorized)	1.00 (0.96-1.05)	.94	1.02 (0.97-1.07)	.53	0.99 (0.94-1.04)	.70	1.00 (0.93-1.07)	.90
Race and ethnicity (reference: white)								
Black	1.71 (0.85-3.42)	.13	0.68 (0.25-1.90)	.47	0.85 (0.32-2.25)	.75	0.76 (0.19-2.99)	.70
Latina	0.95 (0.32-2.83)	.93	0.68 (0.15-3.00)	.61	0.78 (0.17-3.44)	.74	1.20 (0.25-5.73)	.82
Multiracial, Asian, AI/AN[c], NHPI[d], other	1.45 (0.61-3.47)	.40	0.62 (0.14-2.70)	.52	1.74 (0.63-4.81)	.29	1.14 (0.24-5.40)	.87
Education (reference: college degree or higher)								
High school or less	0.59 (0.20-1.74)	.34	0.67 (0.19-2.34)	.53	0.93 (0.27-3.13)	.90	1.27 (0.18-9.22)	.81
Some college	1.28 (0.52-3.15)	.59	1.30 (0.45-3.72)	.62	1.51 (0.51-4.50)	.46	1.80 (0.25-12.71)	.55
Region (reference: South)								
Northeast	0.68 (0.29-1.61)	.38	0.51 (0.17-1.54)	.23	0.60 (0.20-1.80)	.36	Undefined[e]	N/A[f]
Midwest	0.63 (0.31-1.28)	.20)	0.83 (0.39-1.76)	.62	0.94 (0.44-1.99)	.86	0.48 (0.15-1.49)	.20
West	0.82 (0.37-1.80)	.61	0.26 (0.06-1.12)	.07	0.24 (0.05-1.04)	.06	0.57 (0.16-2.11)	.40
Cigarettes per day (reference: light)								
Moderate	1.15 (0.62-2.15)	.66	2.31 (1.12-4.75)	.02	1.40 (0.66-2.96)	.38	2.29 (0.82-6.44)	.12
Heavy	0.59 (0.13-2.58)	.48	1.00 (0.21-4.64)	>.99	0.91 (0.20-4.14)	.90	1.01 (0.12-8.63)	.99
Smoking frequency (reference: nondaily)								
Daily	0.54 (0.26-1.12)	.10	0.58 (0.21-1.62)	.30	0.51 (0.20-1.34)	.17	0.27 (0.08-0.89)	.03
Prequit time	0.98 (0.93-1.03)	.37	1.00 (0.94-1.06)	>.99	1.00 (0.95-1.06)	.94	0.94 (0.87-1.03)	.17

[a]N=1082 reflects total users who made it to quit date. On day 72, n=1053 because 29 users had quit 43-71 days before the end of the study and did not have the opportunity to respond to the day 72 smoking status prompt.

[b]aOR: adjusted odds ratio.

[c]AI/AN: American Indian/Alaska Native.

[d]NHPI: Native Hawaiian and Pacific Islander.

[e]Due to quasi-complete separation of data points, aOR was undefined.

[f]N/A: not applicable.

Discussion

Principal Findings

In a real-world implementation of SmokefreeMOM, an SMS text messaging smoking cessation intervention targeting pregnant women, racial and ethnic composition of users did not mirror the national rates of smoking during pregnancy. Users dropped out as early as the sign-up date and continued to drop out throughout the prequit and intervention periods, with fewer than half completing the intervention. Black women, users with some college education, and users with a high school education or less had a lower likelihood of dropping out of SmokefreeMOM. Nonetheless, response and abstinence rates did not differ by race and ethnicity or by education at key milestones: quit day, intervention end, and 1-month follow-up. Efforts are needed to ensure that SmokefreeMOM reaches and engages pregnant smokers and helps them achieve smoking abstinence, particularly minorities and those with lower educational attainment with comparatively high rates of smoking during pregnancy.

The demographic composition of SmokefreeMOM users revealed adequate enrollment of pregnant smokers of low educational attainment but inadequate enrollment of racial and ethnic minorities. Specifically, women with some college education or high school education or less, who have high rates of smoking during pregnancy (8%-12%) [33], represented 82.68% (1065/1288) of SmokefreeMOM users by educational attainment in our sample. In contrast, AI/AN women, who have the highest rates of smoking during pregnancy by race and ethnicity, as well as NHPI, Asian, and multiracial women, were underrepresented, with AI/AN (9/1230), NHPI (6/1230), and Asian (14/1230) women representing 2.35% and multiracial women representing 4.72% (58/1230) of users. Efforts to overenroll marginalized populations is paramount given their limited access to smoking cessation resources and high risk of exacerbated smoking-related health problems.

Efforts should be directed to improve SmokefreeMOM's overall reach and early uptake among women who intend to become pregnant. Over 4 years, SmokefreeMOM had 2764 user records, meaning that SmokefreeMOM reached less than 1% of roughly a million pregnant smokers (250,000 pregnant smokers per year) [4]. Aside from *how many* women SmokefreeMOM reached, *when* it reached them is equally important. On average, SmokefreeMOM users were in their second trimester, with 5 months until their due dates. A potential approach is to integrate SmokefreeMOM in clinical care for women of reproductive age, as is done with text4baby, an SMS text messaging intervention targeting pregnant women and new mothers with tailored messages about their pregnancy or their babies' development [34]. This integration should include Planned Parenthood and federally qualified health centers to reach underserved women to whom enrolling in SmokefreeMOM solely or in combination with other cessation aids can be recommended [35].

SmokefreeMOM produced somewhat poorer abstinence rates than other smoking cessation interventions. Among SmokefreeMOM users, 1.99% (21/1053) were abstinent at 1 month postintervention. Other SMS text messaging interventions for pregnant women showed 7% to 20% abstinence rates at 1-month follow-up [20,36]. These were incentivized, randomized trials, which may account for the higher abstinence rates. However, abstinence rates among users of SmokefreeTXT, the National Cancer Institute's publicly available SMS text messaging smoking cessation intervention for the general population, were also higher, at 7.2% [37]. Notably, nonresponse for smoking status questions was high and, to be conservative, we considered nonresponders to be smokers. However, nonresponse could signify passive disengagement from the intervention wherein some users failed to text "STOP" to drop out. Conversely, nonresponders or dropouts could be those who quit smoking and had disengaged from messages to avoid smoking relapse triggers [20]. Indeed, 5 of 509 users reported that they were abstinent on the day they dropped out. Increasing responses rates is necessary to accurately estimate abstinence outcomes of SmokefreeMOM.

The lower dropout rates among black users and those with some college education or with high school education or less may be attributed to a greater motivation to quit smoking or lack of access to alternative cessation aids [38]. Conversely, response and abstinence rates were uniform across all user characteristics. Thus, lower dropout among black users and those with less than college education did not translate to better response and abstinence rates. Although program dose has been predictive of abstinence in other SMS text messaging interventions [16], this association varies by many factors (eg, length of program, number of messages received). For example, abstainers in Text2Quit were enrolled longer than nonabstainers, but an overall program dose measure was not associated with smoking abstinence [39]. Lower abstinence rates are consistent with prior research that finds that black and less-educated smokers struggle with smoking abstinence more than white and college-educated smokers do [40].

The need to improve user retention, interaction with SmokefreeMOM until intervention end, and abstinence rates points to engagement as a coveted strategy [41,42]. Engagement strategies could involve incorporating additional keywords the user can text to interact with the intervention or providing congratulatory messages for interaction [43]. Importantly, the best strategy may sometimes be to temporarily provide nothing to the user. At times, not responding to a prompt can signal intervention fatigue, and providing additional content could have negative impacts on user engagement and progress [44]. SMS text messaging interventions must be balanced to engage users but avoid overwhelming them at the same time. Indeed, SMS text messaging interventions that decrease or allow personalized adjustment of message frequency are more effective than those with a fixed frequency [45].

Research is needed to increase SmokefreeMOM's effectiveness, particularly among high-need minorities and marginalized populations. Of interest is fully utilizing mobile technology to deploy race- and ethnicity-specific message libraries that uniquely appeal to their respective target audiences. These types of targeted programs are desired by racial and ethnicity minority populations [18] and are likely to increase engagement and smoking abstinence [46]. Importantly, SMS text messaging interventions can still be inaccessible to low-income women, due to the cost of receiving or sending messages under limited SMS text messaging plans. Accordingly, providing SmokefreeMOM users with credit toward cell phone bills or making text messages to and from intervention services free of charge, currently done by text4baby [34,47], may be a necessary investment. Such proposals are foreseeable given that most health insurance plans cover smoking treatments to varying degrees [48].

Consistent with previous studies [37,41], in this study, daily smoking was a risk factor for poor response and abstinence rates. Longer prequit time was associated with a lower likelihood of dropping out but a lower likelihood of abstinence at quit day. To address the unique needs of daily smokers attempting to quit, a stepped-up approach that incorporates supplemental cessation aids or a tailored message library in its content and delivery schedule, or both, may be needed [49,50]. More information is needed to understand the benefits and downsides to incorporating a prequit period among pregnant women. Although evidence suggests that a planning period prior to quitting is associated with a higher likelihood of abstinence [26,51], this research focused on the general population of smokers. However, pregnant women are more likely to spontaneously quit than other smokers [52] and, thus, their quit attempts may be more successful if implemented immediately.

Strengths and Limitations

This study reflects the strengths and limitations of the real-world implementation of public health interventions. Real-world observational data are invaluable to assess the effectiveness, generalizability, and implementation fidelity of evidence-based interventions in real settings and to inform future experimental studies and trials in a time- and cost-efficient way [53]. SmokefreeMOM users are heterogeneous compared with those enrolled in studies with stringent inclusion criteria, such as randomized controlled trials. Users self-selected to enroll, wanted to quit smoking, and owned a mobile device, presumably

with SMS text messaging plans that allowed receipt of multiple messages a day. This study did not capture technical issues common in SMS text messaging interventions [20] that could have affected study outcomes. We observed high levels of missing data, low response rates (especially as the intervention progressed), and high dropout rates, likely due to a lack of the monetary compensation that is typical of researcher-controlled studies. Low response rates beyond the 1-month follow-up prevented us from assessing postnatal relapse, although as many as 50% of women return to smoking after pregnancy [5], placing their babies at risk of secondhand smoke [6].

We operationalized dropout to include SmokefreeMOM users who texted "STOP" (ie, active dropout) or were unreachable, which could have inflated our dropout rates if unreachable women had reenrolled with a new phone number or had cell phone service stopped due to financial constraints. This operationalization did not capture passive dropouts, that is, users who did not respond to sequential smoking status prompts. Disentangling active and passive dropouts in future work many inform the development and implementation of tailored engagement strategies. Although response and abstinence rates were uniform across racial and ethnic groups in this study, adequate sample sizes could unveil differences in response and abstinence rates previously documented [54]. Future studies can incentivize enrollment and retention to allow for adequate power to assess program outcomes across racial and ethnic groups and socioeconomic gradients.

Conclusions

SMS text messaging interventions are efficacious for smoking cessation [16,45]. SmokefreeMOM, a freely available cessation intervention, is a necessary resource for a hard-to-reach population of pregnant smokers, especially underserved racial and ethnic minorities. Overall abstinence rates among SmokefreeMOM users were lower than among other smoking cessation SMS text messaging interventions. Response and abstinence rates were equivalent across all demographic characteristics of SmokefreeMOM users. Black and less-educated women were more likely to remain in the intervention until its end, presenting opportunities to enhance their engagement and, subsequently, abstinence rates. Research into strategies to increase the reach, engagement, and effectiveness of SmokefreeMOM, particularly for racial and ethnic minority and other marginalized populations with high rates of smoking during pregnancy, is critical for reducing smoking among pregnant women across the United States.

Acknowledgments

Data were analyzed with input from Information Management Services, Inc, an information management services firm, with a specialty in statistical consultation, used by the US National Institute on Minority Health and Health Disparities, Division of Intramural Research (NIMHD DIR).

This research has been funded by the NIMHD DIR, National Institutes of Health (NIH) and the Tobacco Control Research Branch of the National Cancer Institute, NIH, contract #HHSN261201400002B and #HHSN261000011.

Authors' Contributions

KK drafted the manuscript and analyzed data with input from IMS; EG and CR reviewed the manuscript; SE conceptualized the study and critically edited and reviewed the manuscript for intellectual content; and all authors approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
SmokefreeMOM user characteristics, complete case data.
[PDF File (Adobe PDF File)38 KB - mhealth_v7i10e14699_app1.pdf]

Multimedia Appendix 2
Association between user characteristics and dropout.
[PDF File (Adobe PDF File)40 KB - mhealth_v7i10e14699_app2.pdf]

Multimedia Appendix 3
Response and abstinence rates by race and ethnicity, imputed data.
[PDF File (Adobe PDF File)34 KB - mhealth_v7i10e14699_app3.pdf]

Multimedia Appendix 4
Associations between user characteristics and responses rates and abstinence, complete case analysis.
[PDF File (Adobe PDF File)61 KB - mhealth_v7i10e14699_app4.pdf]

References

1. Siu AL. Behavioral and pharmacotherapy interventions for tobacco smoking cessation in adults, including pregnant women: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2015 Sep 22;163(8):622-634. [doi: 10.7326/M15-2023] [Medline: 26389730]
2. Committee on Underserved Women, Committee on Obstetric Practice. Committee Opinion No. 721: Smoking Cessation During Pregnancy. Obstet Gynecol 2017 Dec;130(4):e200-e204. [doi: 10.1097/AOG.0000000000002353] [Medline: 28937573]
3. Pineles BL, Hsu S, Park E, Samet JM. Systematic review and meta-analyses of perinatal death and maternal exposure to tobacco smoke during pregnancy. Am J Epidemiol 2016 Dec 15;184(2):87-97 [FREE Full text] [doi: 10.1093/aje/kwv301] [Medline: 27370789]
4. Martin JA, Hamilton BE, Ostermann MJK, Driscoll AK, Drake P. Births: final data for 2017. Hyattsville, MD: National Center for Health Statistics; 2018. URL: https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_08-508.pdf [accessed 2019-05-10] [WebCite Cache ID 78GgHUPiS]
5. Rockhill KM, Tong VT, Farr SL, Robbins CL, D'Angelo DV, England LJ. Postpartum smoking relapse after quitting during pregnancy: pregnancy risk assessment monitoring system, 2000-2011. J Womens Health (Larchmt) 2016 Dec;25(5):480-488. [doi: 10.1089/jwh.2015.5244] [Medline: 26717489]
6. U.S. Department of Health and Human Services. The health consequences of involuntary exposure to tobacco smoke: a report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2006. URL: https://www.ncbi.nlm.nih.gov/books/NBK44324/pdf/Bookshelf_NBK44324.pdf [accessed 2019-05-10] [WebCite Cache ID 78GguEsqc]
7. Hall ES, Wexelblatt SL, Greenberg JM. Self-reported and laboratory evaluation of late pregnancy nicotine exposure and drugs of abuse. J Perinatol 2016 Dec;36(10):814-818. [doi: 10.1038/jp.2016.100] [Medline: 27388942]
8. Li H, Hansen AR, McGalliard Z, Gover L, Yan F, Zhang J. Trends in smoking and smoking cessation during pregnancy from 1985 to 2014, racial and ethnic disparity observed from multiple national surveys. Matern Child Health J 2018 Dec;22(5):685-693. [doi: 10.1007/s10995-018-2437-x] [Medline: 29397497]
9. Nguyen KH, Subramanian SV, Sorensen G, Tsang K, Wright RJ. Influence of experiences of racial discrimination and ethnic identity on prenatal smoking among urban black and Hispanic women. J Epidemiol Community Health 2012 Apr;66(4):315-321. [doi: 10.1136/jech.2009.107516] [Medline: 20974840]
10. Dominguez TP, Dunkel-Schetter C, Glynn LM, Hobel C, Sandman CA. Racial differences in birth outcomes: the role of general, pregnancy, and racism stress. Health Psychol 2008 Mar;27(2):194-203 [FREE Full text] [doi: 10.1037/0278-6133.27.2.194] [Medline: 18377138]
11. Bryant AS, Worjoloh A, Caughey AB, Washington AE. Racial/ethnic disparities in obstetric outcomes and care: prevalence and determinants. Am J Obstet Gynecol 2010 Apr;202(4):335-343 [FREE Full text] [doi: 10.1016/j.ajog.2009.10.864] [Medline: 20060513]
12. McKinney CO, Hahn-Holbrook J, Chase-Lansdale PL, Ramey SL, Krohn J, Reed-Vance M, Community Child Health Research Network. Racial and ethnic differences in breastfeeding. Pediatrics 2016 Dec;138(2) [FREE Full text] [doi: 10.1542/peds.2015-2388] [Medline: 27405771]
13. Carswell AL, Ward KD, Vander Weg MW, Scarinci IC, Girsch L, Read M, et al. Prospective associations of breastfeeding and smoking cessation among low-income pregnant women. Matern Child Nutr 2018 Dec;14(4):e12622. [doi: 10.1111/mcn.12622] [Medline: 29797420]
14. Coleman T, Chamberlain C, Cooper S, Leonardi-Bee J. Efficacy and safety of nicotine replacement therapy for smoking cessation in pregnancy: systematic review and meta-analysis. Addiction 2011 Jan;106(1):52-61. [doi: 10.1111/j.1360-0443.2010.03179.x] [Medline: 21054620]
15. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
16. 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;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
17. Heminger CL, Schindler-Ruwisch JM, Abroms LC. Smoking cessation support for pregnant women: role of mobile technology. Subst Abuse Rehabil 2016;7:15-26 [FREE Full text] [doi: 10.2147/SAR.S84239] [Medline: 27110146]
18. Borland T, Babayan A, Irfan S, Schwartz R. Exploring the adequacy of smoking cessation support for pregnant and postpartum women. BMC Public Health 2013 May 14;13:472 [FREE Full text] [doi: 10.1186/1471-2458-13-472] [Medline: 23672201]
19. U.S. Department of Health and Human Services, National Cancer Institute. Smokefree.gov. 2019. URL: https://smokefree.gov/ [WebCite Cache ID 78Gh6HsFd]
20. Abroms LC, Chiang S, Macherelli L, Leavitt L, Montgomery M. Assessing the National Cancer Institute's SmokefreeMOM text-messaging program for pregnant smokers: pilot randomized trial. J Med Internet Res 2017 Dec 03;19(10):e333 [FREE Full text] [doi: 10.2196/jmir.8411] [Medline: 28974483]

21. Morehead A, Morse E, Price J. Does the use of motivational text messages and a smoking cessation quitline influence smoking behaviors in pregnant women in Tennessee? J Nurs Patient Care 2017;02(02). [doi: 10.4172/2573-4571.1000116]
22. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
23. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall, Inc; 1986.
24. Stoyneva I, Coa K, Pugatch J, Sanders A, Schwarz M, Cole-Lewis H. SmokefreeTXT behaviour change technique analysis. J Smok Cessat 2016 Dec 20;12(4):231-243. [doi: 10.1017/jsc.2016.24]
25. American College of Obstetricians and Gynecologists. ACOG Committee Opinion No 579: Definition of term pregnancy. Obstet Gynecol 2013 Nov;122(5):1139-1140. [doi: 10.1097/01.AOG.0000437385.88715.4a] [Medline: 24150030]
26. de Vries H, Eggers SM, Bolman C. The role of action planning and plan enactment for smoking cessation. BMC Public Health 2013;13:393 [FREE Full text] [doi: 10.1186/1471-2458-13-393] [Medline: 23622256]
27. Dong Y, Peng CJ. Principled missing data methods for researchers. Springerplus 2013 Dec;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]
28. Little RJA. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc 1988 Dec;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
29. Bennett DA. How can I deal with missing data in my study? Aust N Z J Public Health 2001 Oct;25(5):464-469. [Medline: 11688629]
30. Langkamp DL, Lehman A, Lemeshow S. Techniques for handling missing data in secondary analyses of large surveys. Acad Pediatr 2010;10(3):205-210 [FREE Full text] [doi: 10.1016/j.acap.2010.01.005] [Medline: 20338836]
31. Centers for Disease Control and Prevention. User guide to the 2017 natality public use file. Atlanta, GA: Centers for Disease Control and Prevention; 2017. URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/natality/UserGuide2017.pdf [WebCite Cache ID 74afa9i4l]
32. Ruppert D. Trimming and winsorization. In: Kotz S, Read CB, Balakrishnan N, Vidakovic B, Johnson NL, editors. Encyclopedia of Statistical Sciences. 2nd edition. New York, NY: Wiley; 2006.
33. Drake P, Driscoll AK, Mathews TJ. Cigarette smoking during pregnancy: United States, 2016. Hyattsville, MD: National Center for Health Statistics; 2018. URL: https://www.cdc.gov/nchs/data/databriefs/db305.pdf [accessed 2019-05-10] [WebCite Cache ID 78Gff0374]
34. text4baby. New York, NY: Wellpass; 2019. URL: http://text4baby.org [WebCite Cache ID 78GhNoACI]
35. Glasgow RE, Whitlock EP, Eakin EG, Lichtenstein E. A brief smoking cessation intervention for women in low-income planned parenthood clinics. Am J Public Health 2000 May;90(5):786-789. [doi: 10.2105/ajph.90.5.786] [Medline: 10800431]
36. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction 2017 Jul;112(7):1238-1249 [FREE Full text] [doi: 10.1111/add.13802] [Medline: 28239919]
37. Cole-Lewis H, Augustson E, Sanders A, Schwarz M, Geng Y, Coa K, et al. Analysing user-reported data for enhancement of SmokefreeTXT: a national text message smoking cessation intervention. Tob Control 2017 Nov;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]
38. Borrelli B, Hogan JW, Bock B, Pinto B, Roberts M, Marcus B. Predictors of quitting and dropout among women in a clinic-based smoking cessation program. Psychol Addict Behav 2002 Mar;16(1):22-27. [doi: 10.1037/0893-164X.16.1.22] [Medline: 11934082]
39. Heminger CL, Boal AL, Zumer M, Abroms LC. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Jul;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]
40. Piper ME, Cook JW, Schlam TR, Jorenby DE, Smith SS, Bolt DM, et al. Gender, race, and education differences in abstinence rates among participants in two randomized smoking cessation trials. Nicotine Tob Res 2010 Jun;12(6):647-657 [FREE Full text] [doi: 10.1093/ntr/ntq067] [Medline: 20439385]
41. Coa KI, Wiseman KP, Higgins B, Augustson E. Associations between engagement and outcomes in the SmokefreeTXT program: a growth mixture modeling analysis. Nicotine Tob Res 2018 Apr 16. [doi: 10.1093/ntr/nty073] [Medline: 29668984]
42. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250. [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
43. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
44. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
45. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]

46. Borrelli B. Smoking cessation: next steps for special populations research and innovative treatments. J Consult Clin Psychol 2010 Feb;78(1):1-12. [doi: 10.1037/a0018327] [Medline: 20099945]
47. Whittaker R, Matoff-Stepp S, Meehan J, Kendrick J, Jordan E, Stange P, et al. Text4baby: development and implementation of a national text messaging health information service. Am J Public Health 2012 Dec;102(12):2207-2213. [doi: 10.2105/AJPH.2012.300736] [Medline: 23078509]
48. Centers for Disease Control and Prevention. Coverage for tobacco use cessation treatments. Atlanta, GA: National Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health, Centers for Disease Control and Prevention; 2014. URL: https://www.cdc.gov/tobacco/quit_smoking/cessation/coverage/pdfs/coverage-508-1019.pdf [WebCite Cache ID 78Gf23dvX]
49. Sloan M, Hopewell S, Coleman T, Cooper S, Naughton F. Smoking cessation support by text message during pregnancy: a qualitative study of views and experiences of the MiQuit intervention. Nicotine Tob Res 2017 May 01;19(5):572-577 [FREE Full text] [doi: 10.1093/ntr/ntw241] [Medline: 28403457]
50. Boucher J, Konkle ATM. Understanding inequalities of maternal smoking--bridging the gap with adapted intervention strategies. Int J Environ Res Public Health 2016 Mar 04;13(3) [FREE Full text] [doi: 10.3390/ijerph13030282] [Medline: 26959037]
51. Borland R, Balmford J, Swift E. Effects of encouraging rapid implementation and/or structured planning of quit attempts on smoking cessation outcomes: a randomized controlled trial. Ann Behav Med 2015 Oct;49(5):732-742. [doi: 10.1007/s12160-015-9706-3] [Medline: 25877340]
52. Solomon L, Quinn V. Spontaneous quitting: self-initiated smoking cessation in early pregnancy. Nicotine Tob Res 2004 Apr;6 Suppl 2:S203-S216. [doi: 10.1080/14622200410001669132] [Medline: 15203822]
53. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
54. Robinson CD, Wiseman KP, Hooper MW, El-Toukhy S, Grenen E, Vercammen L, et al. Engagement and Short-term Abstinence Outcomes among Blacks and Whites in the National Cancer Institute's SmokefreeTXT Program. Nicotine Tob Res 2019 Sep 19. [doi: 10.1093/ntr/ntz178] [Medline: 31535690]

Abbreviations

AI/AN: American Indian/Alaska Native
aOR: adjusted odds ratio
HR: hazard ratio
IQR: interquartile range
NHPI: Native Hawaiian and Pacific Islander
SMS: short message service

Edited by G Eysenbach; submitted 13.05.19; peer-reviewed by D Oram, M Jones; comments to author 05.07.19; revised version received 17.07.19; accepted 19.07.19; published 07.10.19.

Please cite as:
Kamke K, Grenen E, Robinson C, El-Toukhy S
Dropout and Abstinence Outcomes in a National Text Messaging Smoking Cessation Intervention for Pregnant Women, SmokefreeMOM: Observational Study
JMIR Mhealth Uhealth 2019;7(10):e14699
URL: https://mhealth.jmir.org/2019/10/e14699
doi:10.2196/14699
PMID:31593542

Original Paper

A Short Message Service Intervention to Support Adherence to Home-Based Strengthening Exercise for People With Knee Osteoarthritis: Intervention Design Applying the Behavior Change Wheel

Rachel K Nelligan[1], BPhysio; Rana S Hinman[1], PhD; Lou Atkins[2], PhD; Kim L Bennell[1], PhD

[1]Centre for Health Exercise and Sports Medicine, Department of Physiotherapy, The University of Melbourne, Parkville, Australia

[2]Centre for Behaviour Change, University College London, London, United Kingdom

Corresponding Author:
Kim L Bennell, PhD
Centre for Health Exercise and Sports Medicine
Department of Physiotherapy
The University of Melbourne
161 Barry Street
Parkville, 3010
Australia
Phone: 61 3 8344 0556
Email: k.bennell@unimelb.edu.au

Abstract

Background: Knee osteoarthritis is a chronic condition with no known cure. Treatment focuses on symptom management, with exercise recommended as a core component by all clinical practice guidelines. However, long-term adherence to exercise is poor among many people with knee osteoarthritis, which limits its capacity to provide sustained symptom relief. To improve exercise outcomes, scalable interventions that facilitate exercise adherence are needed. SMS (short message service) interventions show promise in health behavior change. The Behavior Change Wheel (BCW) is a widely used framework that provides a structured approach to designing behavior change interventions and has been used extensively in health behavior change intervention design.

Objective: The study aimed to describe the development of, and rationale for, an SMS program to support exercise adherence in people with knee osteoarthritis using the BCW framework.

Methods: The intervention was developed in two phases. Phase 1 involved using the BCW to select the target behavior and associated barriers, facilitators, and behavior change techniques (BCTs). Phase 2 involved design of the program functionality and message library. Messages arranged into a 24-week schedule were provided to an external company to be developed into an automated SMS program.

Results: The target behavior was identified as participation in self-directed home-based strengthening exercise 3 times a week for 24 weeks. A total of 13 barriers and 9 facilitators of the behavior and 20 BCTs were selected to use in the intervention. In addition, 198 SMS text messages were developed and organized into a 24-week automated program that functions by prompting users to self-report the number of home exercise sessions completed each week. Users who reported ≥3 exercise sessions/week (adherent) received positive reinforcement messages. Users who reported <3 exercise sessions/week (nonadherent) were asked to select a barrier (from a list of standardized response options) that best explains why they found performing the exercises challenging in the previous week. This automatically triggers an SMS containing a BCT suggestion relevant to overcoming the selected barrier. Users also received BCT messages to facilitate exercise adherence, irrespective of self-reported adherence.

Conclusions: This study demonstrates application of the BCW to guide development of an automated SMS intervention to support exercise adherence in knee osteoarthritis. Future research is needed to assess whether the intervention improves adherence to the prescribed home-based strengthening exercise.

(JMIR Mhealth Uhealth 2019;7(10):e14619) doi:10.2196/14619

KEYWORDS

text messaging; mobile phone; knee osteoarthritis; exercise

Introduction

Knee osteoarthritis is a chronic, highly prevalent condition with no known cure [1] and is a leading contributor to the burden of disease globally [2]. Pain and impaired physical function are characteristic symptoms leading to disability, inactivity, and reduced quality of life [3,4]. Self-management and lifestyle modification to facilitate long-term symptom relief is advocated [5]. This includes exercise, which is recommended in all clinical guidelines irrespective of the person's age, disease severity, pain, physical dysfunction, and/or comorbidities [5-7]. Unfortunately, adherence to exercise is often poor, particularly in the mid to longer term [8-13], limiting its capacity to provide long-term symptom relief. To improve exercise outcomes, scalable interventions that facilitate exercise adherence are a research priority.

Adherence, defined as the extent to which a person's behaviour corresponds with agreed recommendations [14], is complex, multifaceted, and not fully understood [11,15,16]. As such, designing interventions to promote exercise adherence is challenging. To aid complex intervention design, the use of theoretical frameworks is recommended [17-19]. The Behavior Change Wheel (BCW) is a synthesis of 19 models of behavior [20,21] that has been used extensively in intervention design [22-27]. When applying the BCW, the first step of intervention design is to use the Capability, Opportunity, Motivation model of behavior (COM-B) to analyze the desired behavior and identify key barriers to and facilitators of the behavior that the intervention is intended to change. The COM-B model describes three interacting categories that influence behavior: (1) capability that includes physical capability (eg, physical skill) and psychological capability (eg, knowledge and psychological skill); (2) opportunity that includes physical opportunity (eg, the environment such as time and resources) and social opportunity (eg, social cues, norms, and interpersonal influences); and (3) motivation including reflective motivation (eg, self-conscious intentions and beliefs) and automatic motivation (eg, emotional reactions, desires, and impulses). These categories can be further divided into the Theoretical Domains Framework (TDF), 14 additional domains of behavior if greater detail is required [21]. Once the behavior has been analyzed using the COM-B model, the BCW then provides recommendations about the functions that interventions could serve to bring about change (eg, education or training) and guides the selection of potential behavior change techniques (BCTs—*active ingredients* designed to bring about change) that could deliver selected intervention functions.

SMS or mobile phone text messaging programs are becoming an increasingly popular delivery method for health behavior change interventions [28-34]. This is unsurprising considering the widespread use of mobile phones across all populations and age groups [29] and the many benefits of using SMS technology. These include instantaneous communication [30], convenience, cost-effectiveness [35], and, most importantly, high user acceptability [36]. Reviews do, however, suggest the need for caution when drawing conclusions about the effectiveness of SMS health behavior change interventions owing to the overall low quality of individual studies [37,38] and the absence of interventions that are rigorously informed by behavior change theory [34,39].

To our knowledge, only two small studies have specifically assessed mobile phone message interventions to support exercise adherence in people with knee osteoarthritis [40,41]. First, a pilot study on 14 participants evaluated 12 video messages (multimedia messaging service [MMS]), delivered every second day to provide visual prompts to exercise [40]. Over the 6-week intervention, there was no change in functional outcomes or exercise adherence. Participants who received the MMS messages did, however, highly value them as reminders. Second, a feasibility study on 27 participants assessed the effects of an educational booklet about knee osteoarthritis and exercise (delivered by mail) plus 4 weekly activity-promoting text messages [41]. Message content was informed by the social cognitive theory, although how exactly the theory informed the intervention design was not described. Significant improvements in pain and exercise self-efficacy were reported. In total, 96% of participants enjoyed receiving the messages and 88% found them useful. Collectively, the limited research to date suggests that people with knee osteoarthritis find SMS technology an acceptable method to support home exercise.

The ability of automated SMS programs to improve adherence to home-based exercise has been demonstrated in healthy adults [42] and adults with frozen shoulder [43]. A randomized controlled trial (RCT) assessed the effect of a 12-week automated SMS program to promote adherence to weekly home-based strengthening exercise in nonexercising healthy adults (aged 50-70 years) from an upper-middle-income country [42]. Participants in the intervention and control received exercise prescription via a booklet. In addition to the exercise booklet, participants in the SMS intervention also received 60 SMS text messages (1 per weekday) containing BCTs aimed to provide encouragement and exercise reminders. BCT selection was informed by the literature. Messages came from a standard set and were not tailored. At 12 weeks, participants receiving the SMS exercised significantly more than participants who received the exercise booklet only. These benefits were not sustained once SMS contact ceased and had diminished by a 24-week follow-up. A second RCT, which assessed the additive effect of a 2-week automated SMS intervention to therapist-prescribed shoulder exercise for people with frozen shoulder, found significantly higher exercise adherence and improved shoulder range of movement in the SMS group compared with the control group who received exercise instruction only [43]. Messages in this study included reminders, encouragement, and education.

To explore if SMS interventions are effective in promoting exercise adherence in knee osteoarthritis, specifically, theory informed and systematically designed interventions that are clearly reported and rigorously assessed are needed. Therefore, the aim of this study was to describe the development of, and rationale for, an SMS behavior change intervention designed by applying the BCW for people with knee osteoarthritis to support exercise adherence.

Methods

Overview

The SMS intervention was developed in two phases. Phase 1 applied the BCW framework [20,44] to identify barriers to and facilitators of the target behavior to inform selection of the intervention content. Phase 2 developed the SMS program functionality (ie, how users interact with the program) and the SMS text message library, according to published recommendations for the development of mobile phone text message health behavior interventions [19].

Phase 1: Applying the Behavior Change Wheel Framework to Inform Intervention Design

The three stages outlined in the BCW were applied (see Textbox 1) [20,21]. The three stages were initially completed by one author (RN) and then reviewed and discussed by three authors (RN, KB, and RH). The process was overseen by one author (LA), a behavior change expert involved in developing the original BCW framework and with extensive experience in its practical application.

Stage 1 involved gaining a thorough understanding of the behavior. Through reviewing the literature, the problem was defined in behavioral terms and the target behavior selected and explained in detail with context. To identify what needs to change to support the target behavior, we drew on the findings of a scoping review previously conducted by members of our team [45]. This review synthesized key barriers and facilitators for people with hip and/or knee osteoarthritis to participate in intentional exercise, mapped according to the TDF domains. Barriers and facilitators relevant to the target behavior and suitable for SMS delivery were selected from this review and organized according to the COM-B model of behavior.

Stages 2 and 3 used the BCW mapping process to select intervention functions and BCTs to address the barriers and facilitators identified in Stage 1 and with the capacity to be incorporated into the SMS format. First, this involved selecting intervention functions appropriate for SMS delivery that link to the COM-B categories identified in Stage 1. Intervention functions are the types of interventions most likely to bring about change in behavior and include education, persuasion, incentivisation, coercion, training, restriction environmental restructuring, modelling, and enablement. Once intervention functions were selected, BCTs relevant to each intervention function were chosen from the BCT Taxonomy (BCTTv1) [46].

Textbox 1. Stages of intervention design using the behavior change technique.

1. Understand the behavior
    1. Define the problem in behavioral terms
    2. Select target behavior
    3. Specify the target behavior
    4. Identify what needs to change
2. Identify intervention options
    1. Identify intervention functions
    2. Identify policy categories
3. Identify content and implementation options
    1. Identify behavior change techniques
    2. Identify mode of delivery

Phase 2: Development of the SMS Program Functionality and Message Library

Construction of the SMS program functionality (including message type, message frequency, and level of program interaction) was guided by published recommendations [19] and the associated literature. To develop the message library, BCTs linked to the barriers and facilitators identified in phase 1 were converted into an SMS of a maximum 308 characters. A total of 12 people (7 academics working in knee osteoarthritis conservative management, 4 clinical physiotherapists, and 1 person with knee osteoarthritis) individually provided input about message wording. Three authors (RN, RH, and KB) compiled this feedback and used it to construct the final SMS library. Samples of messages were reviewed by one author (LA) to ensure SMS wording accurately reflected the BCW mapping process (from phase 1) and the intended BCT. The final SMS library was then arranged into a 24-week schedule. The literacy demands of all messages were assessed with Web-based readability software (Readable.io, Added Bytes, Ltd), using the Flesch-Kincaid Grade Level, a mathematical calculation to determine the US grade reading level based on word and sentence length. This is a recommended tool to assist health material writing [47] used previously in the literature [48,49]. A SMS practice sequence was also developed to give users the option to practice replying to messages before commencing the 24-week program. The 24-week message schedule and practice sequence were provided to an external company (SMS Solutions PTY LTD) who were responsible for developing the automated text message program and administrator interface. The 24-week

SMS program was trialed by one researcher (RN) and company staff to test functionality and identify message errors.

Results

Phase 1: Applying the Behavior Change Technique Framework to Inform Intervention Design

Stage 1: Understand the Behavior

The problem was identified as lack of adherence to self-directed home exercise. Home-based exercise in knee osteoarthritis typically includes knee strengthening exercise that is prescribed and supervised by a clinician, such as a physiotherapist. After initial supervision, strengthening exercise is then continued unsupervised by a patient in their home. Home programs are reported to have similar symptomatic effects as supervised exercise (individual or group-based) [50,51]; however, adherence to these programs once supervision ceases often declines [16]. To be effective, home-based strengthening programs should follow strength-based protocols [52]. Adherence over the longer term is required to allow muscle adaptation and habit formation to occur. Habit formation, which can take an average of 2 months and up to 8 months, is a key component to the successful adoption of a new behavior [53] and is particularly important as exercise is encouraged on an ongoing basis for individuals with knee osteoarthritis to ensure long-term symptomatic relief [54]. The target behavior was therefore selected to be participation in structured, self-directed progressive home-based strengthening exercises, 3 times a week. Conservatively, a duration of 24 weeks was selected to facilitate habit formation. The target behavior is described in detail in Table 1.

A total of 9 facilitators and 13 barriers to the target behavior were selected from those identified in the extensive scoping review [45] and considered appropriate for an SMS intervention (Textboxes 2 and 3). These were coded using COM-B/TDF domains and listed in BCW intervention mapping tables (Multimedia Appendices 1 and 2). Barriers within the same COM-B category that had a similar meaning were grouped, resulting in 8 barrier categories. COM-B/TDF mapping as outlined by Dobson et al [45] was retained except for barrier *too tired* with all authors identifying this as a reflective motivation barrier in addition to a psychological capability barrier. From the BCW intervention mapping tables, the COM-B categories of psychological capability, reflective motivation, automatic motivation, and social opportunity were identified as the key areas requiring change to support the target behavior. Most barriers were linked to reflective motivation. Most facilitators were linked to physiological capability.

Table 1. Target behavior described in detail.

Target behavior	Participation in a structured, self-directed progressive home-based strengthening exercise, 3 times a week, for 24 weeks
Who needs to perform the behavior?	Individuals with symptomatic knee osteoarthritis
When will they do it?	When convenient to the person with knee osteoarthritis
Where will they do it?	Home-base
How often will they do it?	3 times a week in accordance with exercise guidelines [55] for 24 weeks
With whom will they do it?	Independently

Textbox 2. Facilitators selected as key to participation in structured, progressive strength-based home exercise with the SMS intervention.

Selected facilitators

1. Accurate osteoarthritis disease knowledge
2. Prioritizing exercise
3. Integrating exercise into daily tasks
4. Belief that you are taking control of your own disability
5. Perceived benefits of exercising
6. Belief that exercise is good for health
7. Positive outcome expectations
8. Long- & Short-term goals
9. Receiving medical advice to exercise

Textbox 3. Barriers selected as key to participation in structured, progressive strength-based home exercise with the SMS intervention.

Selected barriers

1. Forgetfulness
2. Too tired
3. Knee pain limiting perceived ability to exercise
4. Concerned exercise is causing pain/damage (includes concern exercise is causing pain + fear of damaging knee further)
5. Lack of improvement with exercises
6. Boredom with exercise (includes lack of enjoyment in exercise + boredom with exercise)
7. Lacking time (includes conflict with routines + lack of time)
8. Life stress (includes family commitments + increased social strain + life events)

Stage 2: Identifying Intervention Options

A total of five intervention functions were selected: (1) education, (2) persuasion, (3) training, (4) environmental restructuring, and (5) enablement. The five selected intervention functions were added to the BCW intervention mapping tables (Multimedia Appendices 1 and 2).

Stage 3: Identifying Content and Implementation Options

To address the 8 barrier categories, 19 BCTs were selected. To address the 9 facilitators, 4 BCTs were selected. Findings from the literature [56-58] were used to reduce the number of possible facilitator BCTs as less facilitator BCT messages were required for the program (refer to Phase 2: Development of how the SMS program functions and the message library). Three BCTs addressed both barriers and facilitators. The BCTs were then added to BCW intervention mapping tables (Multimedia Appendices 1 and 2).

SMS was selected by the authors (KB and RH) as the mode of delivery for an intervention to support exercise adherence at the outset of intervention design. This was based on the available literature identifying SMS as a scalable, effective, efficient, and affordable tool to promote adherence to a range of health behaviors including physical activity and exercise [29-33,59,60]. This decision limited barriers, facilitators, intervention function, and BCT selection to those who would be suitable to be addressed via SMS.

Phase 2: Development of How the SMS Program Functions and the Message Library

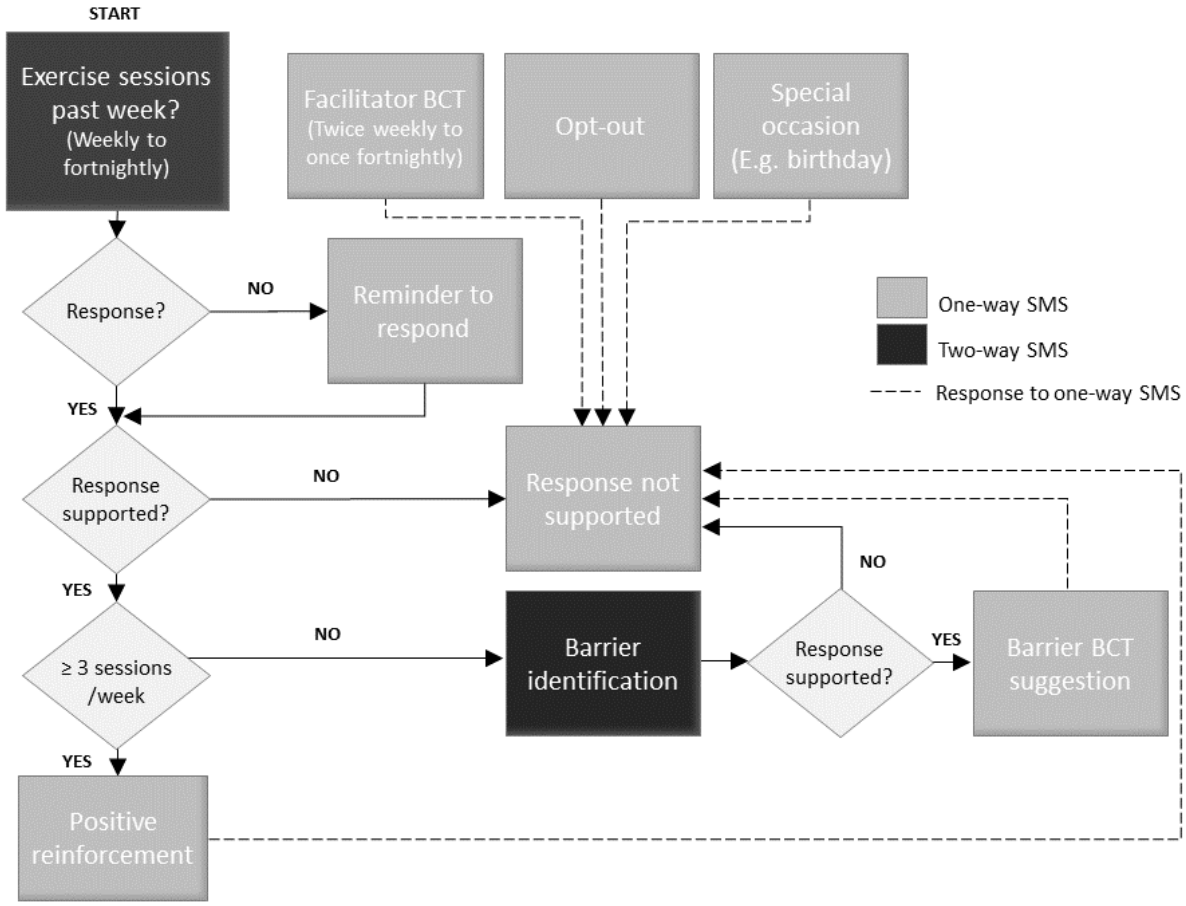
The SMS program was developed to incorporate the barriers and facilitators and BCTs identified in phase 1 and be automated, interactive, and tailored, important characteristics of effective SMS behavior change interventions [19,31].

A total of 9 message types were developed (see Multimedia Appendices 3 and 4) for the automated SMS program. The automated SMS sequence was designed to start with an initial message asking the participant to self-report the number of home exercise sessions they completed in the previous week. Users who self-reported ≥3 exercise sessions/week (adherent) were sent a positive reinforcement message aimed to encourage continued adherence. Users who reported <3 exercise sessions/week (nonadherent) were sent a message asking them to reply by selecting one barrier (from a predefined list of 9 options, including an option of *none above apply to me*) which best explains what made doing the prescribed exercises in the previous week challenging. The user's reply triggers an SMS response containing a BCT suggestion relevant to their selected barrier. BCT messages to facilitate exercise adherence are also sent, irrespective of self-reported adherence. Figure 1 provides a diagrammatic representation of message interactions and triggers.

To address user messages that are not recognized by the program automation (eg, replies to one-way messages or replies not from a predefined list), a *response not supported* message was created encouraging the user to try again or contact program staff if needed. Where appropriate, messages were designed to contain participants' first names. To assist program use (eg, replying to messages), a guided three-message practice sequence was developed. The practice sequence is activated by sending the word START to the phone number assigned to the SMS program.

Guidance from the literature regarding message frequency is highly varied, ranging from daily to once weekly [29,30,33,59], although it has been recommended that 3 messages per week are suitable, with more messages sent initially [19]. Message frequencies that decrease over time appear to be more effective [19,31]. In line with these recommendations, the SMS program was designed to send 4 to 5 messages per week and be reduced to 2 per week by completion of the intervention. Multimedia Appendices 3 and 4 outline message frequency, and the number of each message type required for the 24-week intervention period.

Figure 1. Message interactions and triggers. BCT= behavior change technique.



On the basis of how the program automation works and the message frequency, 198 messages were developed for the intervention: 144 barrier BCT messages (16 per barrier category plus 16 messages relating to *none apply to me*); 24 facilitator BCT messages; 20 positive reinforcement messages; and 10 program logistic messages. A total of 144 barrier messages were required to ensure users do not receive duplicate barrier BCT messages in the event that the same barrier is selected more than once over the 24-week intervention. The literacy demands of all messages in the 24-week schedule were assessed as 5.4 grade, well below the maximum 8th-grade reading level recommended for consumer health care information [19,61]. All messages were arranged into a 24-week schedule, ensuring a spread of BCTs within each COM-B category.

Figure 2 provides an example of an automated message sequence for reported low exercise adherence. Tables 2 and 3 provide examples of the mapping of barrier and facilitator BCT messages following the BCW framework.

The 24-week message schedule and practice sequence was provided to SMS Solutions PTY LTD who were contracted to convert both into an automated text message program with administrative access via a password-protected website. SMS Solutions PTY LTD provided training to the research staff in program and website interface use. The program was then pretested with minor spelling and message duplication errors identified and corrected. The website interface enables addition of program users by entering their name, mobile phone number, birth date, and program start date (set as the following Monday) which triggers the automated 24-week message sequence. The sequence commences with a *Goal Tracking* message. All user messages sent and received are recorded within the program website. All unsupported user communication (eg, replies to one-way messages) is marked with an identifier to allow easy monitoring and subsequent follow-up by research staff if required.

Figure 2. Example automated message sequence for low exercise adherence (<3 exercise sessions) and barrier 'forgot'. BCT= behavior change technique.

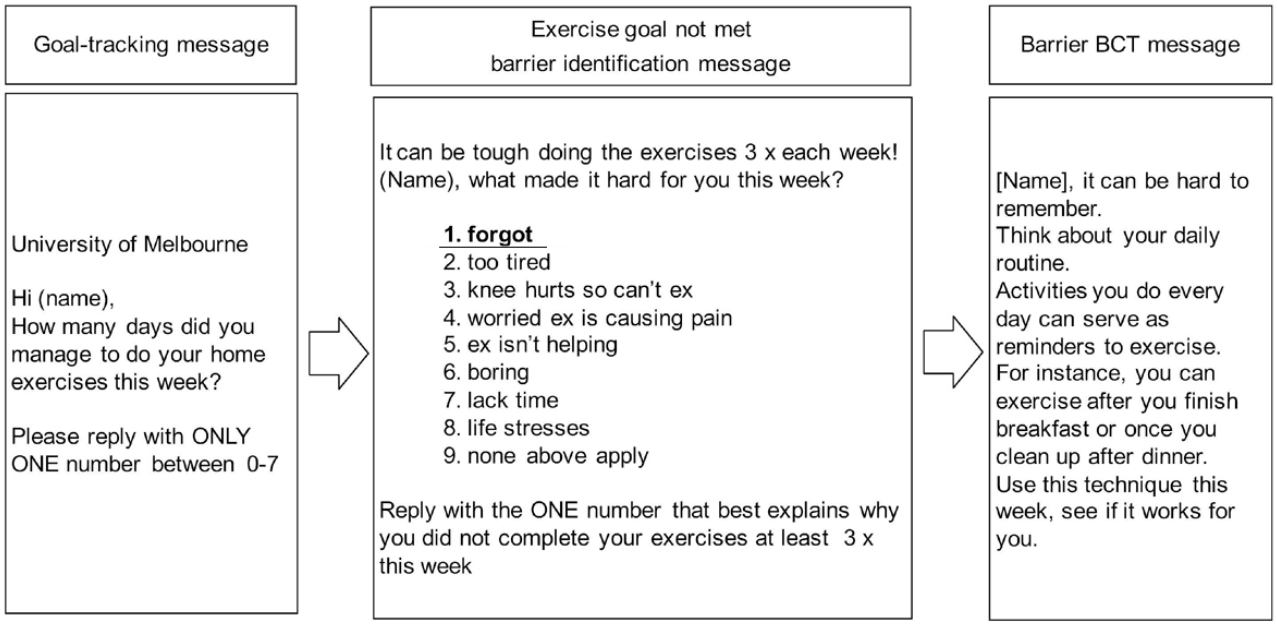


Table 2. Example mapping of barrier behavior change technique messages following the behavior change wheel framework.

Barrier	COM-B[a] category	TDF[b] domain	Intervention function	BCT[c]	SMS[d] content
Forgetfulness	Psychological capability	10. Memory, attention and decision processes	Training	8.3 Habit formation	[Name], it can be hard to remember. We suggest making the exercises a habit. Set aside the same time each day to do them. It's much harder to forget when something is a daily routine.
Concern exercise (causing) pain	Reflective motivation	6. Beliefs about consequences	Enablement	1.2 Problem Solving	Remember mild pain with knee exercise is okay but significant pain's discouraging. Figure out the cause! Use a log book this week, record how you feel after EACH exercise. Identify the exercise linked to your concern. Modify that ONE exercise & keep doing your program 3x week.

[a]COM-B: Capability, Opportunity, Motivation model of behavior.

[b]TDF: Theoretical Domains Framework.

[c]BCT: behavior change technique.

Table 3. Example mapping of facilitator behavior change technique messages following the behavior change wheel framework. OA: osteoarthritis.

Facilitator	COM-B[a] category	TDF[b] domain	Intervention function	BCT[c]	SMS[d] content
Accurate disease knowledge	Psychological capability	1. Knowledge	Education	5.1 Information about health consequences	Let's bust this myth - Surgery is not inevitable if you have knee OA! Exercise is one of the most effective ways to reduce your knee pain and prevent surgery. Stick with your exercise to see the benefits!
Prioritizing exercise	Psychological capability	14. Behavioral regulation	Enablement	10.9 Self reward	Did you prioritize your exercise this week and get them done? Then reward yourself, (name)! Sticking to an exercise program for this long is a real accomplishment that deserves celebration.

[a]COM-B: Capability, Opportunity, Motivation model of behavior.

[b]TDF: Theoretical Domains Framework.

[c]BCT: behavior change technique.

Discussion

Overview

This study reports the systematic design of an intervention to support exercise adherence targeting people with knee osteoarthritis using the BCW framework and current evidence. The intervention is a 24-week automated, semipersonalized mobile phone SMS program to support adherence to prescribed home-based strengthening exercise 3 times a week. The program aims to do this by asking users how many home exercise sessions they completed in the previous week, providing positive reinforcement messages to adherent users, prompting barrier identification and providing targeted behavior change support messages to nonadherent users, and providing facilitation to exercise messages irrespective of weekly exercise adherence.

Strengths and Limitations

The systematic design of the SMS intervention content is a significant strength of this study. Application of the BCW provided a clear framework to develop a targeted and informed intervention that incorporates BCTs specifically selected to address known barriers and facilitators to exercise adherence in people with knee osteoarthritis. The BCW also provided structure for the transparent and thorough reporting of intervention design, vital for future evidence synthesis that will enable a greater understanding of how digital behavior change interventions may have their effect [44]. There are several additional strengths related to the SMS program functionality. Owing to its automated design, minimal administrative support is required throughout the 24-week intervention, making it scalable and potentially cost-effective for both the deliverer and user. The program includes weekly to fortnightly exercise monitoring and provides instantaneous feedback, important characteristics of effective exercise adherence interventions [13,44]. The use of keywords selected from predetermined lists triggers appropriate replies and enables tailoring of content, important in addressing users' individual needs. Intervention fidelity is also guaranteed with all users receiving standardized evidence-based information.

Several limitations should be acknowledged. First, the SMS delivery mode requires users to have adequate vision, a reasonable level of English literacy, access to a mobile phone, and ability to use the SMS function on their phone. However, 87% of Australian adults aged above 55 years use a mobile phone [62] and 75% of all mobile phone users are regular text message senders [63], suggesting this SMS intervention is broadly relevant to most Australians. Second, users who report low exercise adherence must select from a predetermined list of barrier categories that may not specifically target their unique needs. Users are also not able to seek clarification of message content and do not receive additional support that may be required if the same barrier is selected repeatedly. The program functionality does, however, ensure a user will not receive the same BCT message in response to selecting the same barrier. For future implementation of this SMS program, several factors must be considered such as funding sources for intervention delivery (approximately Aus $8 for the 24-week outgoing messages per user) and the need for some level of ongoing administrative support to follow up emergency messages, if received.

Future Research

We intend to evaluate the feasibility, effectiveness, and acceptability of the SMS program to support adherence to prescribed home-based strengthening exercise in RCTs, using both quantitative and qualitative methods. If effective, the intervention could be an easily scalable, cost-effective, convenient solution to support adherence to prescribed home-based strengthening exercise for people with knee osteoarthritis. The intervention could be incorporated into current or future exercise resources (eg, Web-based, remotely delivered exercise programs) or be provided by health professionals to facilitate exercise adherence, improve exercise outcomes, and ultimately improve longer term knee osteoarthritis symptom management. The program could also be adapted for use in other health conditions where exercise adherence is needed.

Conclusions

This study describes how the BCW can be successfully and systematically used to guide development of an automated SMS intervention to support exercise adherence in knee osteoarthritis. Future research is needed to assess whether the intervention improves adherence to prescribed home-based strengthening exercise and is accepted by users, people with knee osteoarthritis.

Acknowledgments

The authors would like to thank SMS Solutions PTY LTD who developed the SMS program platform and online interface. The authors would also like to thank the academics, physiotherapists, and consumers who provided input into SMS wording. This study is funded by the National Health and Medical Research Council (Centre of Research Excellence). RSH is supported by the National Health & Medical Research Council Fellowship (No 1154217). KLB is supported by the National Health & Medical Research Council Fellowship (No 1058440). RKN is supported by the Australian Government Research Training Program Scholarship.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Mapping barrier to exercise to behavior change technique using the Behavior Change Wheel.
[PDF File (Adobe PDF File), 212 KB - mhealth_v7i10e14619_app1.pdf]

Multimedia Appendix 2
Mapping facilitator to exercise to behavior change technique using the Behavior Change Wheel.
[PDF File (Adobe PDF File), 227 KB - mhealth_v7i10e14619_app2.pdf]

Multimedia Appendix 3
Description and frequency of behavior change automated messages.
[PPTX File, 47 KB - mhealth_v7i10e14619_app3.pptx]

Multimedia Appendix 4
Description and frequency of logistic and other automated messages.
[PPTX File, 42 KB - mhealth_v7i10e14619_app4.pptx]

References

1. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol 2014 Jul;10(7):437-441. [doi: 10.1038/nrrheum.2014.44] [Medline: 24662640]
2. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
3. Dieppe P, Cushnaghan J, Tucker M, Browning S, Shepstone L. The Bristol 'OA500 study': progression and impact of the disease after 8 years. Osteoarthritis Cartilage 2000 Mar;8(2):63-68 [FREE Full text] [doi: 10.1053/joca.1999.0272] [Medline: 10772234]
4. Sharma L, Cahue S, Song J, Hayes K, Pai YC, Dunlop D. Physical functioning over three years in knee osteoarthritis: role of psychosocial, local mechanical, and neuromuscular factors. Arthritis Rheum 2003 Dec;48(12):3359-3370 [FREE Full text] [doi: 10.1002/art.11420] [Medline: 14673987]
5. Osteoarthritis: Care and Management - Clinical Guideline. London: National Institute for Health and Care Excellence; 2014.
6. Bannuru RR, Schmid CH, Kent DM, Vaysbrot EE, Wong JB, McAlindon TE. Comparative effectiveness of pharmacologic interventions for knee osteoarthritis: a systematic review and network meta-analysis. Ann Intern Med 2015 Jan 6;162(1):46-54. [doi: 10.7326/M14-1231] [Medline: 25560713]
7. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: the chronic osteoarthritis management initiative of the US bone and joint initiative. Semin Arthritis Rheum 2014 Jun;43(6):701-712. [doi: 10.1016/j.semarthrit.2013.11.012] [Medline: 24387819]
8. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94 [FREE Full text] [doi: 10.1002/acr.22915] [Medline: 27111441]
9. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
10. Pisters MF, Veenhof C, de Bakker DH, Schellevis FG, Dekker J. Behavioural graded activity results in better exercise adherence and more physical activity than usual care in people with osteoarthritis: a cluster-randomised trial. J Physiother 2010;56(1):41-47 [FREE Full text] [doi: 10.1016/S1836-9553(10)70053-9] [Medline: 20500136]
11. Marks R. Knee osteoarthritis and exercise adherence: a review. Curr Aging Sci 2012 Feb;5(1):72-83. [doi: 10.2174/1874612811205010072] [Medline: 21762086]
12. Pisters MF, Veenhof C, van Meeteren NL, Ostelo RW, de Bakker DH, Schellevis FG, et al. Long-term effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review. Arthritis Rheum 2007 Oct 15;57(7):1245-1253 [FREE Full text] [doi: 10.1002/art.23009] [Medline: 17907210]
13. Nicolson PJ, Hinman RS, Kasza J, Bennell KL. Trajectories of adherence to home-based exercise programs among people with knee osteoarthritis. Osteoarthritis Cartilage 2018 Apr;26(4):513-521 [FREE Full text] [doi: 10.1016/j.joca.2018.01.009] [Medline: 29360592]
14. World Health Organization. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organisation; 2003.
15. Jordan JL, Holden MA, Mason EE, Foster NE. Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. Cochrane Database Syst Rev 2010 Jan 20(1):CD005956. [doi: 10.1002/14651858.CD005956.pub2] [Medline: 20091582]

16. Pisters MF, Veenhof C, Schellevis FG, Twisk JW, Dekker J, de Bakker DH. Exercise adherence improving long-term patient outcome in patients with osteoarthritis of the hip and/or knee. Arthritis Care Res (Hoboken) 2010 Aug;62(8):1087-1094 [FREE Full text] [doi: 10.1002/acr.20182] [Medline: 20235201]
17. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
18. Michie S. Designing and implementing behaviour change interventions to improve population health. J Health Serv Res Policy 2008 Oct;13(Suppl 3):64-69. [doi: 10.1258/jhsrp.2008.008014] [Medline: 18806194]
19. Abroms LC, Whittaker R, Free C, van Alstyne JM, Schindler-Ruwisch JM. Developing and pretesting a text messaging program for health behavior change: recommended steps. JMIR Mhealth Uhealth 2015 Dec 21;3(4):e107 [FREE Full text] [doi: 10.2196/mhealth.4917] [Medline: 26690917]
20. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Great Britain: London Silverback Publishing; 2014.
21. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
22. Mc Sharry J, Murphy PJ, Byrne M. Implementing international sexual counselling guidelines in hospital cardiac rehabilitation: development of the CHARMS intervention using the behaviour change wheel. Implement Sci 2016 Oct 10;11(1):134 [FREE Full text] [doi: 10.1186/s13012-016-0493-4] [Medline: 27724957]
23. Fulton EA, Brown KE, Kwah KL, Wild S. StopApp: using the behaviour change wheel to develop an app to increase uptake and attendance at NHS stop smoking services. Healthcare (Basel) 2016 Jun 8;4(2):E31 [FREE Full text] [doi: 10.3390/healthcare4020031] [Medline: 27417619]
24. Barker F, Atkins L, de Lusignan S. Applying the COM-B behaviour model and behaviour change wheel to develop an intervention to improve hearing-aid use in adult auditory rehabilitation. Int J Audiol 2016 Jul;55(Suppl 3):S90-S98. [doi: 10.3109/14992027.2015.1120894] [Medline: 27420547]
25. Sinnott C, Mercer SW, Payne RA, Duerden M, Bradley CP, Byrne M. Improving medication management in multimorbidity: development of the MultimorbiditY COllaborative Medication Review And DEcision Making (MY COMRADE) intervention using the Behaviour Change Wheel. Implement Sci 2015 Sep 24;10:132 [FREE Full text] [doi: 10.1186/s13012-015-0322-1] [Medline: 26404642]
26. Atkins L, Michie S. Designing interventions to change eating behaviours. Proc Nutr Soc 2015 May;74(2):164-170. [doi: 10.1017/S0029665115000075] [Medline: 25998679]
27. 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]
28. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
29. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
30. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12:CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
31. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
32. Chow CK, Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett M, et al. Design and rationale of the tobacco, exercise and diet messages (TEXT ME) trial of a text message-based intervention for ongoing prevention of cardiovascular disease in people with coronary disease: a randomised controlled trial protocol. BMJ Open 2012;2(1):e000606 [FREE Full text] [doi: 10.1136/bmjopen-2011-000606] [Medline: 22267690]
33. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
34. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
35. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mhealth? A systematic review of economic evaluations of mhealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
36. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
37. Vodopivec-Jamsek V, de Jongh T, Gurol-Urganci I, Atun R, Car J. Mobile phone messaging for preventive health care. Cochrane Database Syst Rev 2012 Dec 12;12:CD007457 [FREE Full text] [doi: 10.1002/14651858.CD007457.pub2] [Medline: 23235643]

38. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
39. Ludwig K, Arthur R, Sculthorpe N, Fountain H, Buchan DS. Text messaging interventions for improvement in physical activity and sedentary behavior in youth: systematic review. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e10799 [FREE Full text] [doi: 10.2196/10799] [Medline: 30224335]
40. Dar G, Marx Y, Ioffe E, Kodesh E. The effectiveness of a multimedia messaging service reminder system in the management of knee osteoarthritis: a pilot study. Int J Clin Exp Med 2014;5(9):483-489. [doi: 10.4236/ijcm.2014.59067]
41. Blake H, Roberts AL, Batt ME, Moses JP. Motive8! Feasibility of a text messaging intervention to promote physical activity in knee osteoarthritis. Int J Sports Exerc Med 2015 Dec 31;1(5):1-8. [doi: 10.23937/2469-5718/1510027]
42. Müller AM, Khoo S, Morris T. Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial. J Med Internet Res 2016 Jan 7;18(1):e5 [FREE Full text] [doi: 10.2196/jmir.5235] [Medline: 26742999]
43. Chen HC, Chuang TY, Lin PC, Lin YK, Chuang YH. Effects of messages delivered by mobile phone on increasing compliance with shoulder exercises among patients with a frozen shoulder. J Nurs Scholarsh 2017 Jul;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
44. 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]
45. Dobson F, Bennell KL, French SD, Nicolson PJ, Klaasman RN, Holden MA, et al. Barriers and facilitators to exercise participation in people with hip and/or knee osteoarthritis: synthesis of the literature using behavior change theory. Am J Phys Med Rehabil 2016 May;95(5):372-389. [doi: 10.1097/PHM.0000000000000448] [Medline: 26945211]
46. 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]
47. MedlinePlus. 2016. How to Write Easy-to-Read Health Materials URL: https://medlineplus.gov/etr.html [accessed 2016-02-22]
48. Santos PJ, Daar DA, Badeau A, Leis A. Readability of online materials for Dupuytren's contracture. J Hand Ther 2018;31(4):472-479. [doi: 10.1016/j.jht.2017.07.005] [Medline: 28843342]
49. Saunders CH, Elwyn G, Kirkland K, Durand MA. Serious choices: a protocol for an environmental scan of patient decision aids for seriously ill people at risk of death facing choices about life-sustaining treatments. Patient 2018 Feb;11(1):97-106. [doi: 10.1007/s40271-017-0268-2] [Medline: 28825182]
50. Anwer S, Alghadir A, Brismée JM. Effect of home exercise program in patients with knee osteoarthritis: a systematic review and meta-analysis. J Geriatr Phys Ther 2016;39(1):38-48. [doi: 10.1519/JPT.0000000000000045] [Medline: 25695471]
51. Fransen M, McConnell S, Harmer AR, van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee: a Cochrane systematic review. Br J Sports Med 2015 Dec;49(24):1554-1557. [doi: 10.1136/bjsports-2015-095424] [Medline: 26405113]
52. Bartholdy C, Juhl C, Christensen R, Lund H, Zhang W, Henriksen M. The role of muscle strengthening in exercise therapy for knee osteoarthritis: a systematic review and meta-regression analysis of randomized trials. Semin Arthritis Rheum 2017 Aug;47(1):9-21. [doi: 10.1016/j.semarthrit.2017.03.007] [Medline: 28438380]
53. Lally P, van Jaarsveld CH, Potts HW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]
54. Brosseau L, Taki J, Desjardins B, Thevenot O, Fransen M, Wells GA, et al. The Ottawa panel clinical practice guidelines for the management of knee osteoarthritis. Part two: strengthening exercise programs. Clin Rehabil 2017 May;31(5):596-611. [doi: 10.1177/0269215517691084] [Medline: 28183213]
55. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
56. French DP, Olander EK, Chisholm A, Mc Sharry J. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Ann Behav Med 2014 Oct;48(2):225-234. [doi: 10.1007/s12160-014-9593-z] [Medline: 24648017]
57. Olander EK, Fletcher H, Williams S, Atkinson L, Turner A, French DP. What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2013 Mar 3;10:29 [FREE Full text] [doi: 10.1186/1479-5868-10-29] [Medline: 23452345]
58. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour--and are they the same? Health Educ Res 2011 Apr;26(2):308-322. [doi: 10.1093/her/cyr005] [Medline: 21321008]

59. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
60. Iribarren SJ, Brown W, Giguere R, Stone P, Schnall R, Staggers N, et al. Scoping review and evaluation of SMS/text messaging platforms for mhealth projects or clinical interventions. Int J Med Inform 2017 May;101:28-40 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.017] [Medline: 28347445]
61. Schilling L, Bennett B, Bull S, Kempe A, Wretling MP, Staton E. University of Colorado Denver. 2013. Text Messaging in Healthcare Research Toolkit: The Center for Research in Implementation Science and Prevention (CRISP) URL: http://www.ucdenver.edu/academics/colleges/medicalschool/programs/crisp/training/toolkits/textingtoolkit/Documents/Text%20Messaging%20in%20Healthcare%20Research%20Toolkit%202.pdf [accessed 2019-08-22]
62. Australian Communications and Media Authority. 2009. Use of Digital Media and Communications by Senior Australians URL: https://www.acma.gov.au/theACMA/use-of-digital-media-and-communications-by-senior-australians [accessed 2017-01-20]
63. Kohut A, Wike R, Horowitz JM, Simmons K, Poushter J, Barker C, et al. Pew Research Center. 2011. Global Digital Communication: Texting, Social Networking Popular Worldwide URL: https://www.pewresearch.org/global/2011/12/20/global-digital-communication-texting-social-networking-popular-worldwide/ [accessed 2019-08-22]

Abbreviations

BCT: behavior change technique
BCW: Behavior Change Wheel
COM-B: Capability, Opportunity, Motivation model of behavior
MMS: multimedia messaging service
RCT: randomized controlled trial
TDF: Theoretical Domains Framework

Edited by G Eysenbach; submitted 08.05.19; peer-reviewed by F Barker, T Jiang; comments to author 31.05.19; revised version received 24.07.19; accepted 27.07.19; published 18.10.19.

Please cite as:
Nelligan RK, Hinman RS, Atkins L, Bennell KL
A Short Message Service Intervention to Support Adherence to Home-Based Strengthening Exercise for People With Knee Osteoarthritis: Intervention Design Applying the Behavior Change Wheel
JMIR Mhealth Uhealth 2019;7(10):e14619
URL: https://mhealth.jmir.org/2019/10/e14619
doi:10.2196/14619
PMID:31628786

Original Paper

Health-Related Quality of Life Before and After Use of a Smartphone App for Adolescents and Young Adults With Cancer: Pre-Post Interventional Study

Helle Pappot[1*], MD, DMSc; Gry Assam Taarnhøj[1*], MD; Abbey Elsbernd[2], BSc; Maiken Hjerming[3], RN; Signe Hanghøj[4], MA, MPH, PhD; Marc Jensen; Kirsten Arntz Boisen[4], MD, PhD

[1]Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

[2]University of Kansas School of Medicine, Kansas City, KS, United States

[3]Department of Hematology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

[4]Center of Adolescent Medicine, Department of Pediatric and Adolescent Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

[*]these authors contributed equally

Corresponding Author:
Gry Assam Taarnhøj, MD
Department of Oncology
Rigshospitalet
Copenhagen University Hospital
Blegdamsvej 9
Copenhagen, 2100
Denmark
Phone: 45 51943003
Email: gry.assam.taarnhoej@regionh.dk

Abstract

Background: Adolescent and young adult (AYA) patients with cancer are a group with underexplored needs throughout treatment and in survivorship. This missing knowledge can influence their quality of life (QoL). Given this fact, we have developed a smartphone app based on a cocreation process and have an investigation of QoL among users planned as part of pilot testing this app. Future research is warranted to determine the effect of mobile health (mHealth) tools such as smartphone apps among the AYA cancer population.

Objective: The aim of this study was to investigate the feasibility of a smartphone app among AYA patients with cancer in active treatment and posttreatment, in a pilot test by measuring health-related QoL before and after the use of the app.

Methods: Participants were recruited via the youth support initiative and social organization for AYAs with cancer, Kræftværket, based at Rigshospitalet, University Hospital of Copenhagen, Denmark. Participants were evenly distributed in active treatment and posttreatment groups. After written informed consent, all participants were asked to use the app Kræftværket as they deemed appropriate over a 6-week period. The participants were asked to complete the 30-item European Organization for Research and Treatment of Cancer Quality of Life Questionnaire before and after the 6-week period. The collected QoL data were analyzed with *t* tests to determine differences between groups and from baseline.

Results: In total, we enrolled 20 participants, 10 in active treatment and 10 posttreatment (median time after treatment was 4 months) group. Median age of the participants was 25 years. No differences in QoL were seen at baseline (P=.65). The posttreatment group experienced a significant increase in overall QoL after the 6-week period (global QoL: baseline 62.5, SD 22.3; after 6 weeks 80.8, SD 9.7; P=.04). For the group in active treatment, the QoL remained stable throughout the 6 weeks.

Conclusions: This study shows the feasibility and possible effect on QoL associated with the use of an mHealth tool in AYA patients. mHealth support tools are warranted for this population.

(JMIR Mhealth Uhealth 2019;7(10):e13829) doi:10.2196/13829

KEYWORDS

adolescent; young adult; cancer; mHealth; smartphone; survivorship; quality of life

Introduction

Background

Adolescents and young adults (AYAs) with cancer is a patient group with unique and often underexplored and unmet needs throughout treatment and in survivorship [1-3]. These patients face challenges specific to their age in physical, emotional, and social domains, which may have a detrimental impact on their health-related quality of life (HRQOL) [4-7]. Improving HRQOL in AYA patients with cancer and AYA cancer survivors is of both clinical and research interest in the disciplines of hematology and oncology, particularly through the development of youth-oriented interventions [8].

Smartphone apps and other mobile health (mHealth) interventions have been increasingly used for AYAs with cancer, as well as for AYAs with other acute or chronic disease [9-11]. Apps are attractive tools for AYA oncology and hematology interventions because of their inherent capability of fulfilling a wide number of tasks, including social networking, health tracking, health promotion, and information provision [12,13]. In addition, they are particularly useful for AYA patients, who are often technologically savvy and high users of smartphone technology [13,14]. However, there has been criticism regarding the design of these smartphone apps, based upon limited youth input in design and incomplete or inadequate evaluation of these apps [9,10,15,16].

Despite these limitations, smartphone apps and other mHealth interventions still demonstrate potential to serve as useful interventions for AYAs with cancer and AYA cancer survivors. On the basis of a cocreation process, our research team at Rigshospitalet, University Hospital of Copenhagen, Denmark, has developed a smartphone app, Kræftværket, to improve the QoL in AYAs with cancer and AYA cancer survivors [17,18]. The Kræftværket app, which is named after a youth support initiative and social organization for AYAs with cancer out of Rigshospitalet with the same name, has been designed with 3 primary features: (1) a symptom and activity diary, (2) a communication network between app users, and (3) an information database including video content. All features were selected and refined using the process of cocreation, in which AYA input was used to determine app content and design. The evaluation protocol for this app has been designed according to 2 evaluation stages: pilot testing and implementation testing [19]. The pilot testing consists of a qualitative and quantitative launch of the app to a small group of 20 AYA patients with cancer and cancer survivors, with quantitative evaluation using the 30-item European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) instrument, and qualitative evaluation using focus group interviews and think-aloud testing [19]. Results of pilot testing will be used to modify and improve the app for launch of the app to a larger number of participants for implementation testing.

Objectives

The aim of this study was to present the quantitative HRQOL data from the Kræftværket app pilot testing.

Methods

Overview

Detailed methods on the Kræftværket app development project are described in 2 separate papers [17,19]. As mentioned previously, the Kræftværket app was designed with 3 primary features: (1) a symptom and activity diary, (2) a communication network between app users, and (3) an information database including video content. The communication network allowed for nonmediated group-based communication between users as requested by the users during the cocreation process. The study plan was to evaluate the feasibility of the use of the Kræftværket app while measuring QoL in a pre-post study design.

Participants and Recruitment

Kræftværket is a youth support initiative and social organization for AYA patients with cancer aged 15 to 29 years based at Rigshospitalet University Hospital of Copenhagen, Denmark [17]. Inclusion criteria for app pilot testing were AYAs aged 15 to 29 years with prior Kræftværket initiative contact and access to a smartphone and the internet, including cellular data or Wi-Fi. Participants were excluded if they were unable to read and write in Danish, and if they participated in the cocreation activities that determined app content and design [19]. For the pilot-testing phase, 20 previous or current Kræftværket users were recruited to pilot-test the app and provide qualitative and quantitative feedback. Recruitment was targeted to obtain 10 participants currently undergoing treatment for cancer, and 10 participants who had completed treatment for cancer. Participants were invited to participate in the study through invitation in the open Facebook group Kræftværket Rigshospitalet or while participating in other Kræftværket initiatives at Rigshospitalet. No individual patient was approached. The demographic data were provided by the AYAs themselves.

Pilot Testing

After obtaining informed consent, participants provided baseline measurements of HRQOL using the validated EORTC QLQ-C30 instrument for overall and subdomain QoL measurement. No other EORTC modules or instruments were used throughout this pilot testing.

Participants were asked to download the Kræftværket app and contact the software developer, Daman, to obtain a log-in, followed by using the Kræftværket app over the course of 6 weeks. Upon completing written informed consent, participants were shown the app and the 3 features by youth coordinator MH, but they were not given any specific instructions on suggested frequency of app use; instead, they were asked to use the app as they deemed fit. At the end of the 6-week period, they were asked to complete a secondary EORTC QLQ-C30. The original protocol for app evaluation stated that participants would be prompted to complete the EORTC QLQ-C30 via the app; however, this instrument was completed in paper form because of changes in personal data protection rules.

At this time, additional qualitative data were collected by author SH in the form of semistructured focus group interviews and individual think-aloud tests. The qualitative data were collected

after the completion of the QLQ-C30 after the 6-week period. Data from the qualitative aspects of this pilot test will be published in 2 separate papers [20,21].

Statistical Analysis

Statistical analysis was performed with IBM SPSS version 25.0. Descriptive statistics and frequencies were used to display sociodemographic and clinical data. A paired *t* test was performed to determine differences in global QoL from baseline. Differences over time are illustrated by a boxplot. No statistical analyses were performed between groups as the sample size did not allow for between-group comparisons.

Ethical Considerations

We obtained informed consent from all individual participants included in the study, whereas caregiver informed consent was given for patients younger than 18 years. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This research was exempt from review by an institutional review board or ethical authority under Danish law. Due to the Danish Data Protection law restrictions, the tracking data registered through the app of the patients participating in this study were not available. Permission to conduct the study was granted by the Data Protection Agency (j.nb. 2012-58-0004, i-suite nb.:6217).

Results

Participants and Recruitment

Over a 2-week period, 21 patients were approached at Kræftværket; of these, only 1 patient declined participation. In total, 20 recruited patients, 10 in active treatment and 10 posttreatment, completed written informed consent. Of these participants, 70% (14/20) were female. Among the participants in the active treatment group, the gender distribution was 60% (6/10) vs 40% (4/10) with a majority of females. This dispersion increased in the posttreatment group with 80% (8/10) women. Median age of the participants was 25 years, and for the posttreatment group, the median time elapsed from diagnosis to start of study was 4 months. The type of cancer was predominantly hematologic cancer (50%, 10/20) followed by breast cancer (20%, 4/20). Table 1 presents the demographic and clinical data.

Pilot Testing

The mean global QoL at baseline was similar in the 2 treatment groups (active treatment group 66.67 vs posttreatment group: 62.5, *P*=.65); see boxplot in Figure 1. Table 2 shows the average EORTC QLQ-C30 domain scores. All domains are scored on a 0 to 100 scale. Although a higher global QoL score indicates better QoL, the symptom scales, for example, pain and nausea, are reversely scored, with a higher score indicating more impairment. These scores reveal a difference in baseline scores in favor of the posttreatment group for the subdomains role functioning, emotional functioning, cognitive functioning, and social functioning, and the single items for fatigue, nausea, dyspnea, insomnia, appetite loss, and financial difficulties. On the contrary, the single items on pain, constipation, and diarrhea were higher at baseline in the posttreatment group. Due to the large disparity between diagnoses and treatments received, no statistical analyses were performed for these data. A significant increase in global QoL was found for the posttreatment group from baseline to 6 weeks (difference estimate 18.3; 95% CI 1.5-35.1; *P*=.04). The paired analysis found no difference from baseline to 6 weeks after the use of the app for the group in active treatment (difference estimate 1.7; 95% CI –5.6 to 9.0; *P*=.61).

Table 1. Demographic and clinical information of participants (N=20).

Clinical data	All participants (N=20)	Active treatment group (n=10)	Posttreatment group (n=10)
Gender, n (%)			
Male	6 (30)	4 (40)	2 (20)
Female	14 (70)	6 (60)	8 (80)
Age (years), mean (range)	25 (16-29)	24 (19-29)	28 (16-29)
Cancer type, n (%)			
Lymphoma	9 (45)	5 (56)	4 (44)
Breast	4 (20)	3 (75)	1 (25)
Head and neck	2 (10)	1 (50)	1 (50)
Leukemia	1 (5)	0 (0)	1 (100)
Testicular	1 (5)	0 (0)	1 (100)
Ventricular	1 (5)	0 (0)	1 (100)
Thyroid	1 (5)	1 (100)	0 (0)
Brain	1 (5)	0 (0)	1 (100)
Median time posttreatment, months (range)	—[a]	—	4 (1-41)

[a]Not applicable.

Figure 1. Boxplot of difference in baseline and 6-week global quality of life scores.

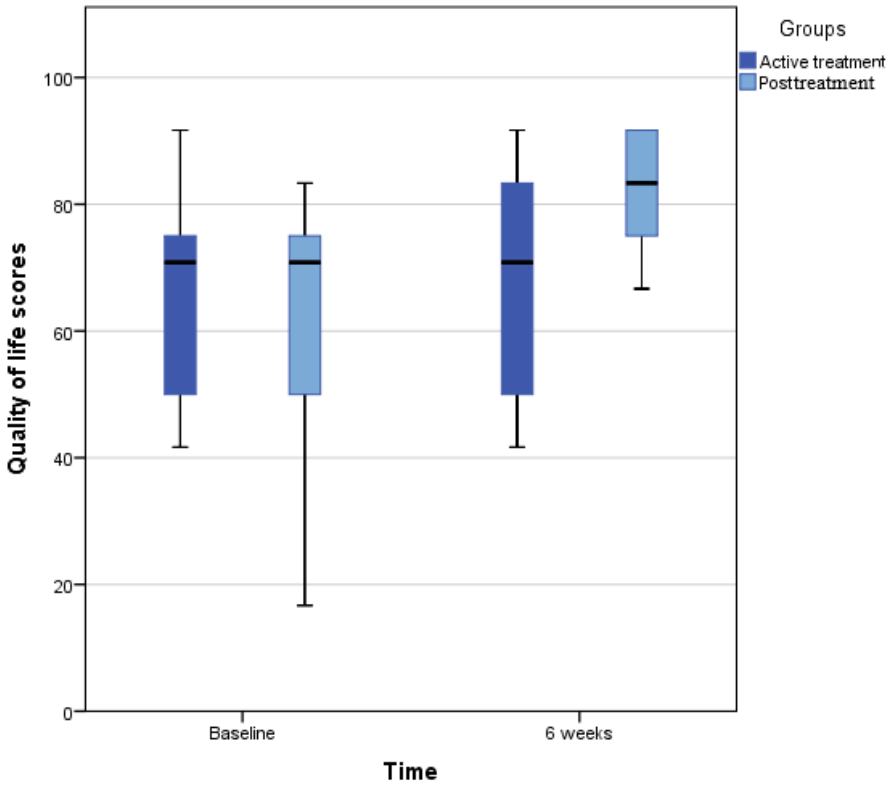


Table 2. Mean value for the 30-item European Organization for Research and Treatment of Cancer Quality of Life Questionnaire scales at baseline and 6-week measurement.

EORTC QLQ-C30[a] domains	Active treatment, mean (SD)		Posttreatment, mean (SD)	
	Baseline	6 weeks	Baseline	6 weeks
Global health status/quality of life	66.67 (17.12)	68.33 (17.48)	62.5 (22.31)	80.83 (9.66)
Physical functioning	74.67 (30.27)	77.33 (30.66)	80.67 (15.85)	87.33 (15.85)
Role functioning	55 (32.44)	60 (27.44)	68.33 (30.88)	81.67 (32.82)
Emotional functioning	65.83 (27.34)	63.33 (21.59)	68.67 (28.47)	82.67 (12.25)
Cognitive functioning	55 (36.89)	71.67 (33.38)	63.33 (33.15)	86.67 (17.21)
Social functioning	73.33 (30.63)	75 (25.15)	78.33 (30.48)	81.67 (21.44)
Fatigue	50 (21.11)	44.44 (23.42)	42.22 (22.71)	33.33 (13.86)
Nausea	20 (31.23)	23.33 (31.62)	10 (11.65)	11.67 (13.72)
Pain	22 (23.64)	25 (30.68)	33.33 (34.25)	13 (15.32)
Dyspnea	33.33 (27.22)	23.33 (31.62)	16.67 (17.57)	6.67 (14.05)
Insomnia	50 (28.33)	36.67 (39.91)	30 (33.15)	30 (29.19)
Appetite loss	30 (36.68)	26.67 (37.84)	10 (16.10)	6.67 (14.05)
Constipation	6.67 (14.05)	13.33 (32.20)	16.67 (32.39)	16.67 (23.57)
Diarrhea	10 (22.50)	13.33 (32.20)	13.33 (23.31)	6.67 (14.05)
Financial difficulties	50 (39.28)	50 (39.28)	13.33 (23.31)	3.33 (10.54)

[a]EORTC QLQ-C30: 30-item European Organization for Research and Treatment of Cancer Quality of Life Questionnaire.

Discussion

Principal Findings

This study has demonstrated feasibility and possible impact of a smartphone app on the HRQOL in this small AYA population. Our results show that QoL may be positively affected in a group of patients posttreatment after a short period of use of a specifically designed mHealth tool.

The field of mHealth technologies designed for individuals with cancer is in a period of rapid growth and development. The present smartphone app may fill the gap for a population in need of added support. Other mHealth tools may have the same potential but have until now lacked testing and validation in an appropriate population [9,10,15]. Several studies have affirmed the feasibility of mHealth use in different populations, thus creating opportunities for similar strategies in other cancer populations, such as older adults, in both active treatment and survivorship, as well as AYA with other conditions.

The increase seen in QoL may be explained by several factors. Posttreatment patients not in daily contact with fellow patients or hospital personnel may experience improvement in QoL using the Kræftværket app owing to added social interaction and increased sense of belonging. A study by Kaal et al introduces the relationship between empowerment and HRQOL in AYA patients and demonstrates that empowerment is positively associated with HRQOL [22]. Empowerment has previously been theorized as a broad construct of intrapersonal, interactional, and behavioral components [23]. When extending these constructs to our data, the interactional component (eg, the social interaction feature of the app) may be able to explain the development seen, although, in this study, we do not have data on which of the 3 individual elements of the app were used most frequently by participants. Thus, we cannot be sure if the increase in QoL is a result of increased empowerment through an element of social interaction. Patients still in active treatment have regular contact with their health care provider, and, in our case, the Kræftværket group and associated personnel. As such, they may not experience an added sense of belonging through using an interactive app such as Kræftværket. This may be an explanation for this group of participants not experiencing the same benefits as the posttreatment group. However, for some AYA populations, for example, AYAs with hematological cancers going through burdensome treatments, one may expect a decrease in QoL. In this study, 50% of the AYAs in active treatment were treated for hematological cancer, yet no decrease in QoL was seen for this group. Due to the study design, the course of QoL for patients not exposed to the app cannot be reported. This can only be explored in a randomized trial design. However, literature informs us of the many issues presumed to affect the QoL of these patients after treatment. Therefore, one would expect a decrease in QoL after treatment [24]. This decrease is not seen in this study, thus emphasizing the possible effect of the Kræftværket app.

The levels of global QoL recorded at baseline by EORTC QLQ-C30 are comparable with previous published data for posttreatment AYA patients [25,26], although lower than the levels in an adult cancer population [27]. The difference in age compared with the adult cancer populations may have an impact on QoL because younger people in important developmental phases of life have higher expectations to life and, thus, a worse experience of the misfortunes of life [28]. In addition, an explanation for the lower overall QoL could be the dominant group of hematologic cancer patients in this and similar AYA studies [25,26]. These patients go through a longer and by many accounts more demanding oncological treatment. Interestingly, the 6-week global QoL value for the posttreatment group in this study was comparable with an age-standardized group of German citizens without cancer as shown by Geue et al [25], thus demonstrating a potential for the Kræftværket app in this population.

Given the potential of the current app, and similar apps for an AYA population in need of added support, the access to smartphone technology is a requirement. A review from 2015 indicates a concern for individuals of lower sociodemographic status and restricted access to smartphones as high as 25% [10]. Restricted access would undoubtedly create a selection bias of whom would benefit from this technology. However, several surveys performed in 2015-2018 from comparable countries have shown very high access to smartphone for the age group of 12 to 34 years (89%-98%) [14,29,30].

Strengths of this study include the feasibility testing of a smartphone app for patients with cancer including the possible positive effects its use has on these patients. Moreover, this app has been tested in a diverse AYA population with a large variety of cancer diseases represented. In addition, the app is tested in 2 different settings, during active treatment and posttreatment, thus implying its use in a survivorship setting for the potential benefit of a large population. Finally, the cocreation process during the development of this app and during the process of this study with direct patient involvement (coauthor MJ) may certainly have a role in the success of its use, hence encouraging a similar process in the development of future apps [19].

Some limitations do need to be addressed. First, the pre-post study design in a very small population clearly represents a limitation and lessens the possible interpretations of the presented data. In addition, owing to the data protection restrictions in Denmark, the tracking information of app use was not available to report. Therefore, we do not know how and to what extent the participants used the app, thereby rendering it difficult to conclude why the posttreatment group experienced an increase in QoL, which may have been seen even without the use of the app. In addition, because the clinical information of cancer disease was self-reported by the patients, we did not have data on the stage of disease or specific oncological treatment. These data may have contributed to a more thorough explanation of the observed lower global QoL scores in these participants, as patients at the end of their lives would expect to experience a decrease in QoL [22]. Finally, although the EORTC QLQ-C30 has been used in more than a thousand clinical trials and daily clinics, several groups have allegated that this instrument is noncomprehensive and does not cover all the QoL aspects important to AYAs [4-6]. In previous studies by Nightingale et al and Quinn et al, 3 constructs were described as lacking for the AYA population: perceived sense of self,

relationships, and parenthood [31,32]. Therefore, the QoL results may not portray the complete picture of these patients' QoL.

Perspectives

This study shows the feasibility and possible positive effects on QoL by use of an mHealth tool in AYA patients. The study gives hope to the use of mHealth tools to improve QoL in AYAs with cancer. Future studies evaluating the impact of the Kræftværket app in the AYA cancer population will show the effect implementation of mHealth tools may have on this patient group.

Acknowledgments

The study was funded by Trygfonden (grant number 120451). The authors highly appreciate the participation of AYAs with cancer associated with Kræftværket. The authors also acknowledge the scientific discussions with members of the Kræftværket steering group.

Conflicts of Interest

None declared.

References

1. Epelman CL. The adolescent and young adult with cancer: state of the art -- psychosocial aspects. Curr Oncol Rep 2013 Aug;15(4):325-331. [doi: 10.1007/s11912-013-0324-6] [Medline: 23700032]
2. Saloustros E, Stark DP, Michailidou K, Mountzios G, Brugieres L, Peccatori FA, et al. The care of adolescents and young adults with cancer: results of the ESMO/SIOPE survey. ESMO Open 2017;2(4):e000252 [FREE Full text] [doi: 10.1136/esmoopen-2017-000252] [Medline: 29018578]
3. Galán S, de la Vega R, Pires CT, Racine M, Solé E, Jensen MP, et al. What are the needs of adolescents and young adults after a cancer treatment? A Delphi study. Eur J Cancer Care (Engl) 2017 Mar;26(2):E12488. [doi: 10.1111/ecc.12488] [Medline: 27007727]
4. Quinn GP, Gonçalves V, Sehovic I, Bowman ML, Reed DR. Quality of life in adolescent and young adult cancer patients: a systematic review of the literature. Patient Relat Outcome Meas 2015;6:19-51 [FREE Full text] [doi: 10.2147/PROM.S51658] [Medline: 25733941]
5. Sodergren SC, Husson O, Robinson J, Rohde GE, Tomaszewska IM, Vivat B, EORTC Quality of Life Group. Systematic review of the health-related quality of life issues facing adolescents and young adults with cancer. Qual Life Res 2017 Jul;26(7):1659-1672 [FREE Full text] [doi: 10.1007/s11136-017-1520-x] [Medline: 28251543]
6. Sodergren SC, Husson O, Rohde GE, Tomasewska IM, Vivat B, Yarom N, et al. A life put on pause: an exploration of the health-related quality of life issues relevant to adolescents and young adults with cancer. J Adolesc Young Adult Oncol 2018 Aug;7(4):453-464. [doi: 10.1089/jayao.2017.0110] [Medline: 29565709]
7. Badr H, Chandra J, Paxton RJ, Ater JL, Urbauer D, Cruz CS, et al. Health-related quality of life, lifestyle behaviors, and intervention preferences of survivors of childhood cancer. J Cancer Surviv 2013 Dec;7(4):523-534 [FREE Full text] [doi: 10.1007/s11764-013-0289-3] [Medline: 23749663]
8. Marris S, Morgan S, Stark D. 'Listening to patients': what is the value of age-appropriate care to teenagers and young adults with cancer? Eur J Cancer Care (Engl) 2011 Mar;20(2):145-151. [doi: 10.1111/j.1365-2354.2010.01186.x] [Medline: 20477855]
9. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
10. Wesley KM, Fizur PJ. A review of mobile applications to help adolescent and young adult cancer patients. Adolesc Health Med Ther 2015;6:141-148 [FREE Full text] [doi: 10.2147/AHMT.S69209] [Medline: 26316835]
11. Castensøe-Seidenfaden P, Husted GR, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a self-management app for young people with type 1 diabetes: methodological challenges, experiences, and recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]
12. Fiordelli M, Diviani N, Schulz PJ. Mapping mhealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
13. BinDhim NF, Trevena L. There's an app for that: a guide for healthcare practitioners and researchers on smartphone technology. Online J Public Health Inform 2015;7(2):e218 [FREE Full text] [doi: 10.5210/ojphi.v7i2.5522] [Medline: 26392848]
14. Anderson M, Jiang J. Pew Research Center. 2018. Teens, Social Media & Technology 2018 URL: http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2019-08-27]
15. Bender JL, Yue RY, 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]

16. Yasini M, Beranger J, Desmarais P, Perez L, Marchand G. mHealth quality: a process to seal the qualified mobile health apps. Stud Health Technol Inform 2016;228:205-209. [doi: 10.3233/978-1-61499-678-1-205] [Medline: 27577372]
17. Elsbernd A, Boisen KA, Hjerming M, Niemann CU, Petersen G, Pappot H, et al. Developing age-appropriate supportive facilities, resources, and activities for adolescents and young adults with cancer across departments and diagnoses: a single-center experience. J Adolesc Young Adult Oncol 2019 Mar;8(1):98-102. [doi: 10.1089/jayao.2018.0077] [Medline: 30394824]
18. Elsbernd A, Hjerming M, Visler C, Hjalgrim LL, Niemann CU, Boisen KA, et al. Using cocreation in the process of designing a smartphone app for adolescents and young adults with cancer: prototype development study. JMIR Form Res 2018 Nov 1;2(2):e23 [FREE Full text] [doi: 10.2196/formative.9842] [Medline: 30684439]
19. Elsbernd A, Hjerming M, Visler C, Hjalgrim LL, Niemann CU, Boisen K, et al. Cocreated smartphone app to improve the quality of life of adolescents and young adults with cancer (Kræftværket): protocol for a quantitative and qualitative evaluation. JMIR Res Protoc 2018 May 10;7(5):e10098 [FREE Full text] [doi: 10.2196/10098] [Medline: 29748162]
20. Hanghøj S, Boisen KA, Hjerming M, Elsbernd A, Pappot H. Think-aloud testing of a smartphone app aimed at adolescents and young adults in and after cancer treatment. JMIR Cancer 2019 (forthcoming). [doi: 10.2196/preprints.15008]
21. Hanghøj S, Boisen KA, Hjerming M, Pappot H. Adolescents and young adults' everyday life experiences of a cancer smartphone app. JMIR Mhealth Uhealth 2019 (forthcoming). [doi: 10.2196/preprints.13829]
22. Kaal SE, Husson O, van Duivenboden S, Jansen R, Manten-Horst E, Servaes P, et al. Empowerment in adolescents and young adults with cancer: relationship with health-related quality of life. Cancer 2017 Oct 15;123(20):4039-4047 [FREE Full text] [doi: 10.1002/cncr.30827] [Medline: 28696580]
23. Zimmerman MA. Psychological empowerment: issues and illustrations. Am J Community Psychol 1995 Oct;23(5):581-599. [doi: 10.1007/BF02506983] [Medline: 8851341]
24. Lea S, Martins A, Bassett M, Cable M, Doig G, Fern LA, et al. Issues experienced and support provided to adolescents and young adults at the end of active treatment for cancer: a rapid review of the literature. Eur J Cancer Care (Engl) 2018 Nov;27(6):e12972. [doi: 10.1111/ecc.12972] [Medline: 30485604]
25. Geue K, Sender A, Schmidt R, Richter D, Hinz A, Schulte T, et al. Gender-specific quality of life after cancer in young adulthood: a comparison with the general population. Qual Life Res 2014 May;23(4):1377-1386. [doi: 10.1007/s11136-013-0559-6] [Medline: 24197479]
26. Tremolada M, Bonichini S, Basso G, Pillon M. Perceived social support and health-related quality of life in AYA cancer survivors and controls. Psychooncology 2016 Dec;25(12):1408-1417. [doi: 10.1002/pon.4072] [Medline: 26810123]
27. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Mar 15;22(4):714-724. [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
28. Eiser C, Penn A, Katz E, Barr R. Psychosocial issues and quality of life. Semin Oncol 2009 Jun;36(3):275-280. [doi: 10.1053/j.seminoncol.2009.03.005] [Medline: 19460584]
29. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social media use and access to digital technology in US young adults in 2016. J Med Internet Res 2017 Jun 7;19(6):e196 [FREE Full text] [doi: 10.2196/jmir.7303] [Medline: 28592394]
30. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
31. Nightingale CL, Quinn GP, Shenkman EA, Curbow BA, Zebrack BJ, Krull KR, et al. Health-related quality of life of young adult survivors of childhood cancer: a review of qualitative studies. J Adolesc Young Adult Oncol 2011 Sep;1(3):124-132 [FREE Full text] [doi: 10.1089/jayao.2011.0033] [Medline: 23610733]
32. Quinn GP, Huang IC, Murphy D, Zidonik-Eddelton K, Krull KR. Missing content from health-related quality of life instruments: interviews with young adult survivors of childhood cancer. Qual Life Res 2013 Mar;22(1):111-118 [FREE Full text] [doi: 10.1007/s11136-012-0120-z] [Medline: 22286223]

Abbreviations

AYA: adolescent and young adult
EORTC QLQ-C30: 30-item European Organization for Research and Treatment of Cancer Quality of Life Questionnaire
HRQOL: health-related quality of life
mHealth: mobile health
QoL: quality of life

Edited by G Eysenbach; submitted 26.02.19; peer-reviewed by X Skrabal Ross, K Absolom, J Bender; comments to author 02.05.19; revised version received 06.06.19; accepted 21.07.19; published 03.10.19.

Please cite as:
Pappot H, Assam Taarnhøj G, Elsbernd A, Hjerming M, Hanghøj S, Jensen M, Boisen KA
Health-Related Quality of Life Before and After Use of a Smartphone App for Adolescents and Young Adults With Cancer: Pre-Post Interventional Study
JMIR Mhealth Uhealth 2019;7(10):e13829
URL: https://mhealth.jmir.org/2019/10/e13829
doi:10.2196/13829
PMID:31584008

Original Paper

Quality of Deaf and Hard-of-Hearing Mobile Apps: Evaluation Using the Mobile App Rating Scale (MARS) With Additional Criteria From a Content Expert

Ryan Lee Romero[1*], BPH; Frederick Kates[1*], PhD, MBA; Mark Hart[1*], EdD; Amanda Ojeda[1], BS; Itai Meirom[1], BS; Stephen Hardy[1*], MEd

College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States

[*]these authors contributed equally

Corresponding Author:
Ryan Lee Romero, BPH
College of Public Health and Health Professions
University of Florida
1225 Center Drive
Gainesville, FL
United States
Phone: 1 (352) 273 6060
Email: ryan.romero1000@ufl.edu

Abstract

Background: The spread of technology and dissemination of knowledge across the World Wide Web has prompted the development of apps for American Sign Language (ASL) translation, interpretation, and syntax recognition. There is limited literature regarding the quality, effectiveness, and appropriateness of mobile health (mHealth) apps for the deaf and hard-of-hearing (DHOH) that pose to aid the DHOH in their everyday communication and activities. Other than the star-rating system with minimal comments regarding quality, the evaluation metrics used to rate mobile apps are commonly subjective.

Objective: This study aimed to evaluate the quality and effectiveness of DHOH apps using a standardized scale. In addition, it also aimed to identify content-specific criteria to improve the evaluation process by using a content expert, and to use the content expert to more accurately evaluate apps and features supporting the DHOH.

Methods: A list of potential apps for evaluation was generated after a preliminary screening for apps related to the DHOH. Inclusion and exclusion criteria were developed to refine the master list of apps. The study modified a standardized rating scale with additional content-specific criteria applicable to the DHOH population for app evaluation. This was accomplished by including a DHOH content expert in the design of content-specific criteria.

Results: The results indicate a clear distinction in Mobile App Rating Scale (MARS) scores among apps within the study's three app categories: ASL translators (highest score=3.72), speech-to-text (highest score=3.6), and hard-of-hearing assistants (highest score=3.90). Of the 217 apps obtained from the search criteria, 21 apps met the inclusion and exclusion criteria. Furthermore, the limited consideration for measures specific to the target population along with a high app turnover rate suggests opportunities for improved app effectiveness and evaluation.

Conclusions: As more mHealth apps enter the market for the DHOH population, more criteria-based evaluation is needed to ensure the safety and appropriateness of the apps for the intended users. Evaluation of population-specific mHealth apps can benefit from content-specific measurement criteria developed by a content expert in the field.

(JMIR Mhealth Uhealth 2019;7(10):e14198) doi:10.2196/14198

KEYWORDS

eHealth; mobile health; mHealth; mobile app; hearing; deaf persons; sign language

Introduction

Background

The continuous growth of mobile phone technology opens new opportunities for mobile health (mHealth) apps. The improved computing capability, increased memory, and the open operating systems of smartphones support mHealth app development and help shape the future of health care [1,2]. In 2018, there were approximately 205.4 billion mobile apps downloaded worldwide, with a forecasted growth to 258.2 billion by 2022 [3]. This emerging technology can create a more inclusive and accessible environment for the deaf and hard-of-hearing (DHOH) population, representing more than 7 million Americans or over 2% of the US population [4,5]. New mHealth apps leveraging the internet can help to reduce barriers for individuals with disabilities, which is a crucial component of the Americans with Disabilities Act [6]. Disability and health inclusion strategies include identifying and eliminating communication barriers for people with hearing impairments [7]. Universal accessibility of apps that can increase the multidirectional communication between DHOH persons and those who are not is essential for social inclusion [8].

Although there are mHealth apps available to the DHOH population, there is minimal information regarding the quality, features, effectiveness, and maintenance of these apps. Information regarding the status of DHOH apps should be expanded; this information will give consumers valuable knowledge relevant to choosing an app. In addition, information on specific user needs could assist developers in recognizing features desired by the population that have not been fulfilled with the present mobile apps. Relying on star ratings and reviews may be insufficient for app developers and analysts because of the volume of ratings and the usefulness of the information [9]. To evaluate and rank apps relevant to the DHOH in a quantitative manner, a standardized scale is needed, such as the Mobile App Rating Scale (MARS) by Stoyanov et al [10]. The MARS is an evaluation system divided into 5 core sections that can accommodate a wide variety of mHealth apps.

The deaf population can be divided into groups based on different criteria such as the degree of hearing loss, language preference, educational experience, and integration in the Deaf community or the hearing population [11]. Throughout, the study will use DHOH to indicate the DHOH population in the United States. Within this population, there is a distinction to be made between persons identifying as Deaf (uppercase D) and deaf (lowercase D). Those who identify as Deaf are actively engaged in a common Deaf culture and the identity behind the culture and prefer to use American Sign Language (ASL) or use only ASL [12]. Persons identifying as deaf or hard-of-hearing may comprise those who have postlingual hearing loss, prefer to use English over ASL, or choose to associate with the hearing culture [13]. Owing to the unique characteristics of the DHOH population in the United States and the complexity of the ASL [14], there is no commonly used written system for ASL; Web-based text is needed for updating content or simple user queries [15]. In the United States, more than 500,000 individuals use sign language as their primary mode of communication [16]. ASL interpreters are commonly employed to resolve communication barriers between the DHOH and others [17]. The Bureau of Labor Statistics projects 18% growth for the industry from 2016 to 2026, which is much faster than the average for all occupations. The United States is expected to add 12,100 new positions by 2026 [18]. However, it can take years to become fluent in ASL [19], and in many areas of the country, there is a shortage of sign language interpreters [17,20]. Although signing interpreters are a staple of interpersonal communication for the DHOH community, mobile apps can also bolster interpersonal communication in numerous ways [21,22].

Objective

The aim of this study was to modify a standardized scale to evaluate and rank apps designed for the DHOH. The study sought the use of a content expert to develop content-specific criteria that would gauge the presence of features relevant to the target population.

Methods

Search and Selection Criteria

An initial screening of the DHOH apps was conducted in May 2018 across 2 mobile app stores: Apple's App Store and the Google Play Store (for Android platform). The initial search criteria sought apps aimed toward assisting the DHOH community in the United States.

The following search terms were used: "deaf," "deaf application," "deaf hearing," "hard of hearing," "sign language translator," "sign language," "sign language applications," "ASL translator," ASL sign language," "sign language dictionary," "sign language keyboard," "text to speech," "hearing assistant," "hearing aid," "deaf text to speech," "deaf translator," and "ASL assistant."

A list of potential apps to review was generated after a preliminary screening was done based on the app description in the stores and supporting screenshots to establish relevance and initial inclusion. In-store reviews and star ratings were not considered to prevent rater bias. The study was IRB201802567 exempt.

Inclusion Criteria of Apps and Processes

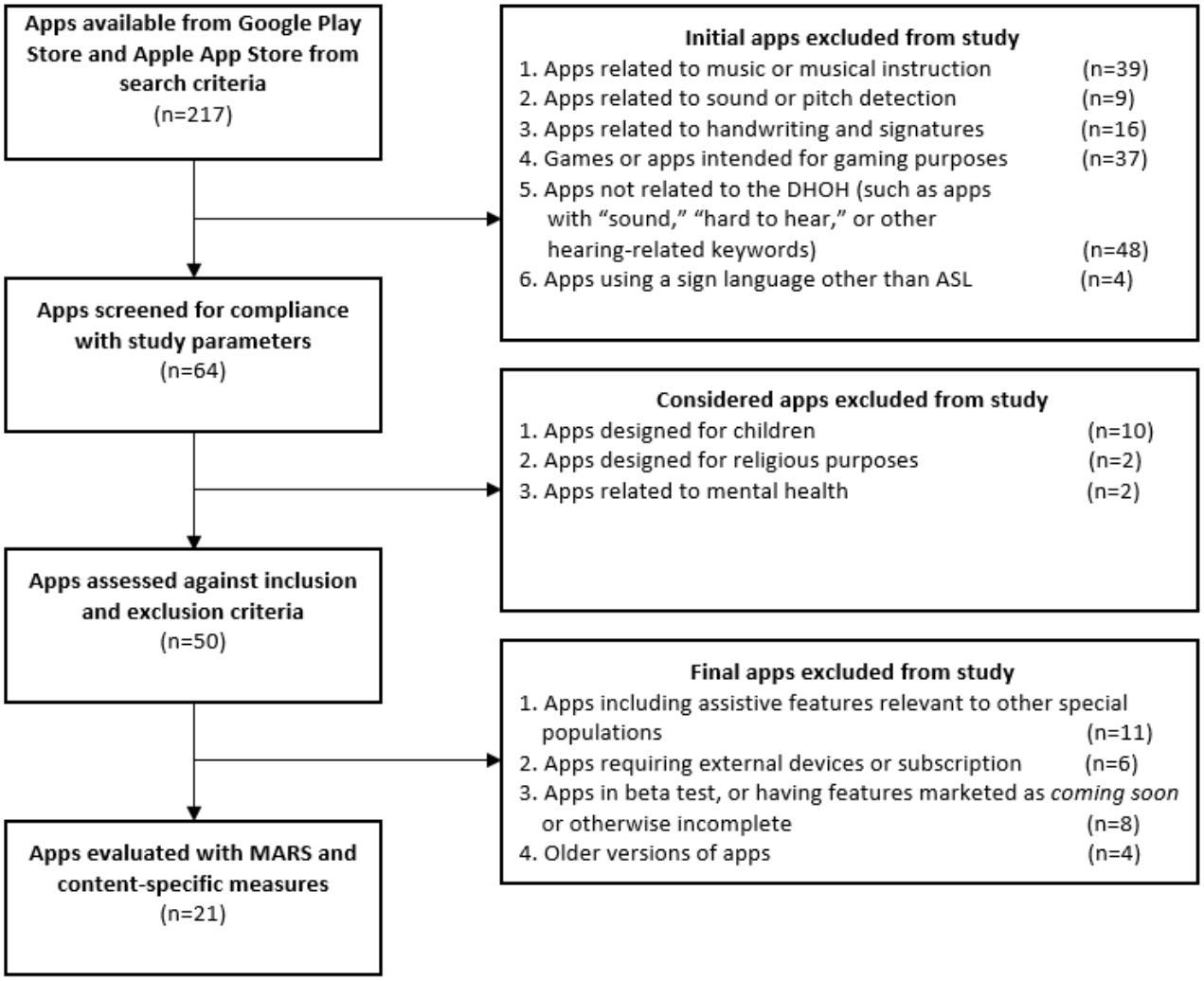
To best select a list of apps, inclusion and exclusion criteria were developed for the master list of apps with the goals of the study in mind, which were (1) to determine the objective quality of hard-of-hearing apps on the mobile market and (2) to gauge the affinity of these apps to integrate the hard-of-hearing population into the national community via the elimination of a social barrier. As the 2 main resource pools for target apps were the Apple's App Store and the Google Play Store, only apps from these 2 repositories were considered.

Initial Apps Excluded From the Study

Apps from both stores (n=217) underwent an initial filtering to obtain a diverse spread of apps related to the DHOH community based on the search criteria. Apps not directly related to the DHOH were not considered further, which include apps related

to music or musical instruction (n=39); apps related to sound or pitch detection (n=9); apps related to handwriting or signatures (n=16); games or apps intended for gaming purposes (n=37); apps not otherwise related to the DHOH, or contained keywords related to DHOH (such as apps with "sound," "hearing," "hard to hear," or other hearing-related keywords within the app title) but did not include DHOH features (n=48); and apps using a sign language other than ASL (eg, British Sign Language or French Sign Language; n=4; see Figure 1).

Figure 1. Inclusion and exclusion criteria flowchart. ASL: American Sign Language; DHOH: deaf and hard-of-hearing; MARS: Mobile App Rating Scale.



Considered Apps Excluded From the Study

After the initial filtering, apps were screened for compatibility with the 3 categories of the study (ASL translators, speech-to-text, and hearing assistants; n=64). Apps having a primary focus other than the aid of the DHOH population were excluded in this step. Apps designed for children (n=10), religious purposes (n=2), or mental health topics (n=2) were not considered for inclusion criteria screening (see Figure 1).

Final Apps Excluded From the Study

Remaining apps (n=50) were then assessed against the inclusion and exclusion criteria. Apps meeting 4 final exclusion criteria were not considered for evaluation with the MARS: apps including assistive features relevant to other special populations (apps with assistive features intended for other populations were excluded to standardize the scoring of the study's content-specific measures; n=11); apps requiring external devices or a subscription for function (n=6); apps in beta test, or having features such as *coming soon* or otherwise incomplete (n=8); and apps representing an older version of an app being considered (n=4; see Figure 1).

The use of mHealth apps typically falls under the definition of assistive technology (AT), which can be a piece of equipment, software program, or product used to increase, maintain, or improve the functional capabilities of persons with disabilities [23]. Therefore, the study includes a content expert (SH) to evaluate the MARS criteria and to create criteria to better match particular assistive apps to specific needs. Owing to the complexity of evaluating ASL apps, particularly for a person who has no hearing difficulties or ASL experience, it was evident that a content expert was needed.

Apps

The development of a multicategory master list of apps allowed for an independent evaluation of each category of apps with the

MARS scale and developed content-specific measures (see Figure 2).

Figure 2. Master list of apps. ASL: American Sign Language.

ASL translators	
App	Device
Signily	iPhone/iPad/Android
The ASL App	iPhone/iPad/Android
ASL Provider	iPhone/ iPad
APEX VRI	Android
Interpret Live	iPhone/iPad/Android
Fluency Mobile	iPhone/iPad
ASL Coach	Android

Hard-of-hearing assistants	
App	Device
Visual Hearing Aid	iPhone/iPad/Android
Petralex Hearing Aid	iPhone/iPad/Android
ListenClear	iPhone/iPad
BeWarned-App for Deaf and Hard of Hearing	iPhone/iPad
Earfy	iPhone/iPad/Android
eyeHear	iPhone/iPad
Sound Alert	iPhone/iPad/Android
Live Caption	iPhone/iPad/Android

Speech-to-text apps	
App	Device
Text to Speech!	iPhone/iPad
Speak4Me-Text to Speech	iPhone/iPad/Android
TextToSpeech	iPhone/iPad
Transcribe-Speech to Text	iPhone/iPad
Text to Speech for iMessage	iPhone/iPad
Smart Record: Audio Recorder	iPhone/iPad/Android

American Sign Language Translators

Apps providing translational functionality to and from ASL were considered in this section. Statistically, "congenital hearing loss affects two to three infants per 1,000 live births" [24]. As ASL is the main component of communication for many Deaf persons in the United States [25], in addition to the unique accessibility needs of Deaf persons regarding communication [11], it was appropriate to allocate an entire section of study to this app category. ASL is unique from verbal language as it provides a mode of communication based on symbols and visual cues [25]. In addition, those who are prelingually deafened (and who typically associate with the Deaf culture) are more reliant on ASL-based communication than verbal or written English [26]. Apps that serve to translate sign language to and from

words, whether it be by visual interpretation or ASL dictionary function, were sorted into this category.

Speech-to-Text

Threaded speech-to-text apps are also helping individuals with hearing loss to take part in conversations with family and groups of people. The best hard-of-hearing lip-readers can understand approximately 30% of a dialogue [27]. With threaded speech-to-text apps, the accuracy of conversation increases to around 80% to 90% [28]. Speech-to-text allows for a more conversational-speed communication between the hard-of-hearing and others. In the current app market, there are a substantial number of these apps competing to be used, yet many are not geared toward the hard-of-hearing population. Although this category of apps poses great benefit to those who were postlingually deafened or are hard of hearing, speech-to-text translators may not be particularly useful toward individuals who were prelingually deafened, stemming from a lower level of English usage and comprehension [29].

Hearing Assistants

The study's third category included apps that sought to improve user communication and interaction in everyday life. This section was created to give the study the opportunity to analyze apps that did not fit into the other 2 categories yet provided some feature that positively affected a DHOH person's ability to communicate. A content expert was used to determine whether or not an app provided some benefit to a DHOH person in social function or public navigation.

Data Collection and the Mobile App Rating Scale

Decision was made to use a reliable and flexible app quality rating scale designed by expert research panelists to assess the quality of mHealth apps via multiple descriptive factors [10]. The MARS was developed to satisfy the need for a reliable and objective scale that could rate the degree to which mHealth apps satisfy the defined quality criteria [10]. The study used the 4 MARS classifications, plus an additional *content-specific* section: engagement, functionality, aesthetics, and information quality. The subjective quality section of the MARS was substituted for the custom-designed content-specific measures section to better gauge app quality and appropriateness. With the inclusion of the study's 4 custom-designed criteria, apps were evaluated on 23 metrics. All the metrics were quantified by assigning integer values: 1=poor, 2=fair, 3=acceptable, 4=good, and 5=excellent.

The mobile apps using the MARS were analyzed by 3 members of the research team (RR, IM, and AO), designated as *raters*. Each rater underwent 2 training sessions to correctly attribute scores from the MARS to the apps. These training sessions were supplemented by a video developed by the MARS creators to aid in the rater training and calibration [30]. The MARS training video is a reference created by the authors of the MARS to explain the purpose of the scale, the app characteristics that each subsection of the scale measures, and the guidelines for evaluating each subsection. The video describes the relative quality of each app characteristic that is appropriate for each score (poor-excellent), along with examples of quality features. Furthermore, the MARS training video acted as a knowledge base during the construction of content-specific measures and generated ideas as to how content-specific criteria may be scored. These ideas aided the refinement of the content-specific questions after development with the content expert. The 3 raters are members of the College of Public Health who are involved in hard-of-hearing research. The raters were all trained with knowledge from the content expert to give a shared and equal understanding of the needs and characteristics of the DHOH population. A total of 3 control apps were used to determine the intraclass correlation coefficients among raters based on Shrout and Fleiss' guidelines [31]. A test similar to Fleiss' kappa, the Krippendorff alpha, was used to test for interrater agreement [32]. An alpha of .91 was obtained. All issues regarding agreeability were discussed among team members and reevaluated for appropriate concurrence.

Use of a Content Expert and "Content-Specific" Measures

Content experts can be useful when determining the relevance of a proposed study to a specific population or subject [33]. Specifically, the study revolves around the needs, desires, and attributes of the DHOH population. The study's content expert (SH) is a professor of ASL who himself is deaf. The Deaf culture is a broad network of beliefs, values, and rules for behaviors that are incorporated into the lives of many with hearing difficulties [12]. The use of a content expert allowed the study design to consider the needs and desires of this population as related to mobile app features so that valid and meaningful metrics could be designed to test for the presence of such features.

An additional scoring section, the content-specific section, was tailored with a content expert (SH) for hearing-specific apps following the MARS authors' interest in criteria applicable to specific populations [10]. This section included 4 subfactors: signing space, distractions, assistive features, and societal integration potential. To help create the content-specific classifications and improve validity, a content expert with experience in teaching ASL and ASL linguistics and Deaf studies was added to the study. After the initial review of apps, concerns were discussed with the content expert to further refine the criteria. All the category factors were quantified by assigning integer values: 1=poor, 2=fair, 3=acceptable, 4=good, and 5=excellent.

Signing Space

Evaluating the signing space is an important measurement criterion as this area gauges the digital interpreter's use of sign language. In sign language, the signing space encompasses the distinctive locations surrounding the signer [34], particularly the space in front of the signer extending from the waist to the forehead [35]. The content expert emphasized that signing in apps should be in a consistent space and easy to follow. The content expert developed several questions with measurable criteria to assist the evaluators in scoring apps: Does the signing stay within approximately 12 inches of the body's center mass? Does the reader have to concentrate on entirely different areas when facial expressions are used? Is the signing too fast or too slow?

Distractions

Distraction evaluation includes any attribute or behavior by the app or interpreter that may be distracting to the reader. These include distracting clothing (with patterns or multiple colors), bright nail polish or jewelry, and extraneous events occurring in the background or the background color [15,36]. Furthermore, signers should consider the contrast between their skin complexion and their clothing and the background to focus the attention around the hands [37].

Assistive Features

The initial purpose of hard-of-hearing apps and translators is to allow those with hearing difficulties to achieve social function similar to those who are not hard of hearing. One way the apps can achieve this is by integrating the assistive features within their menus and the graphical user interfaces (GUIs). These assistive features can be (but are not limited to) enlarged buttons and text, a magnifier function, closed captioning on app dialogue, and slow scrolling text of a large font style.

Societal Integration

One desirable function of apps that seek to assist the DHOH and similar populations would be to provide functions that further aid in the user's inclusion and experience in society. The nature of certain apps may lead them to have a higher affinity for this criterion, but overall, this section can be rated by reviewing the potential for a particular app to integrate a DHOH person into society (via social outings, dating, or simply acting as an aid in public).

Results

App Turnover

One outcome not expected was the high level of *app turnover* noted in the ASL translator section. For the study's purposes, *app turnover* has been defined as an app being added to the master list of apps for evaluation but later being wholly inaccessible or unusable because of its removal from software repositories, extremely poor design, or abandonment by the developer. A total of 5 of the 7 apps were lost to app turnover: ASL Provider, APEX VRI, Interpret Live, Fluency Mobile, and ASL Coach. Although the study was left with a reasonable distribution of apps to evaluate, it should be noted that the high level of app turnover serves as a general reflection of the state of hard-of-hearing apps.

Scoring for American Sign Language Translators

Upon completion of scoring for the ASL translator apps, the consensus among the 3 evaluators was that this category of apps had quality apps, yet it was subject to a large amount of *app turnover*, as previously discussed. Of the 7 ASL translator apps selected for evaluation, only 2 were accessible or even discoverable in either app repositories. The best app rated in the category was Signily (3.72/5), with the second best being The ASL App (3.642/5). Both apps had acceptable to good general attributes and acceptable content-specific qualities (see Table 1).

Scoring for Speech-to-Text

Of the 6 apps tested in the speech-to-text category, all 6 apps were still accessible at the end of the scoring period. The best app in the category was Text to Speech! by Gwyn Durbridge (3.595/5), followed by Speak4Me–Text to Speech (3.584/5) and Smart Record: Audio Recorder (3.48/5). This category of apps had the lowest average score for all 4 general attribute categories, and the lowest average score in the content-specific measures section (see Table 1). However, because these apps were not explicitly designed for the DHOH population, they should only be faulted if drastic improvements toward the target population's needs would be easily achievable.

Scoring for Hearing Assistants

Of the 8 apps tested in the hearing assistant category, all 8 apps were still accessible at the end of the scoring period. The best app in the hearing assistant category was ListenClear (3.90/5), followed by Petralex Hearing Aid (3.89/5) and Sound Alert (3.73). Of the 3 categories evaluated, consensus was reached that this particular category boasted the most consistent level of engagement across apps (see Table 1). In addition, the average functionality score was the highest in this category. Although the quality of the general attributes of these apps is either acceptable or good, several of the apps had only fair or poor scores for the content-specific measures (see Table 1). There was a notable variance among raters in the content-specific section while evaluating the app *ListenClear*. Although the ListenClear app has an exceptional GUI, assistive features, and minimal distractions, it would be desirable for this app to have increased societal integration features considering how it is marketed to the DHOH population (as a hearing assistant).

Table 1. Mobile App Rating Scale app quality ratings.

Mobile health app ranking	Engagement	Functionality	Aesthetics	Information quality	Content-specific criteria[a]
ASL[b] translators					
Signily	3.8	3.8	3.2	3.8	3.9
The ASL App	3.0	4.4	3.9	3.8	3.0
DHOH[c] assistants					
ListenClear	4.1	4.3	4.7	4.0	2.5
Petralex Hearing Aid	3.9	4.3	3.9	3.9	3.4
Sound Alert	3.8	4.3	4.1	3.5	3.0
BeWarned–App for DHOH	4.2	4.1	3.7	3.6	3.1
Visual Hearing Aid	3.3	4.1	3.2	3.6	2.0
eyeHear	2.7	3.8	3.0	3.1	2.2
Live Caption	2.8	3.8	2.6	2.9	2.4
Earfy	2.9	3.3	2.0	2.4	1.8
Speech-to-text					
Text to Speech!	3.9	3.8	3.6	3.5	3.2
Speak4Me–Text to Speech	3.9	4.1	3.4	3.4	3.0
Smart Record: Audio Recorder	3.1	3.8	4.1	4.4	1.0
Transcribe–Speech to Text	3.1	3.3	3.4	3.2	2.1
TextToSpeech (Iconic Solutions, LLC)	2.6	2.8	2.7	2.4	1.5
Text to Speech for iMessage	1.9	1.7	1.9	2.1	1.3

[a]Content-specific criteria are based on the interest in an *app-specific* section by the original Mobile App Rating Scale authors to evaluate apps for specific populations.

[b]ASL: American Sign Language.

[c]DHOH: deaf and hard-of-hearing.

Discussion

The State of Hard-of-Hearing Assistance Technology

There is both a need and demand for continuing development of Android and iOS app support for the DHOH. This study's findings agree with previous findings that app developers prefer Android and iOS for their projects [38]. Of the 217 apps from the search criteria, 50 apps were assessed against the inclusion and exclusion criteria for the study. Given the relatively low yield of candidate apps from the App Store and Google Play Store, it is thought necessary for more development of apps targeted toward the DHOH population. A refinement of the master list of apps to 21 apps based on the inclusion and exclusion criteria was thought sufficient to mitigate any anomalies in the study. When it came time to evaluate the apps, however, it was observed that 5 of the 7 hard-of-hearing assistants had become unavailable or had otherwise disappeared from the app stores. This suggests a high rate of app turnover in the subject field, which may contribute to the limited availability of quality DHOH apps. It may be inferred that DHOH apps are experiencing high turnover rates, consistent with the current AT turnover rates as high as 75% to 80% [39,40]. It seems apparent that although these apps may have been designed by skilled developers, features that may be basic necessities to some persons in the DHOH population are being overlooked or poorly implemented. Furthermore, the relative lack of high-level features, such as an app's ability to act as a social advocate for the user, underlines the state of the DHOH apps: available, but limited in scope.

The Value of Hard-of-Hearing Assistance Technology

One of the tenets of the study was to determine the quality of the available DHOH apps so that it may be inferred on their value to DHOH persons. Given the current state of mobile assistance technology for the intended population, it is reasonable to say that these apps provide augmentation to a deaf or hard-of-hearing person's ability to navigate in public, interact both publicly and with family, and be connected to other deaf persons through the availability of features that facilitate interpersonal interaction. It is notable, however, that there should be more research into other purposes for these apps. For example, there may be a useful purpose for these kinds of apps in both domestic and emergency scenarios such as hurricanes or severe storms. Indeed, the development of emergency AT could help to reduce the disproportionately high level of morbidity among the DHOH during natural disasters [41]. Mobile apps or devices could have a preprogrammed functionality that assists DHOH persons in an emergency, although this warrants further research.

Use of a Content Expert to Evaluate Mobile Apps

One measure seeking to validate the study's design and metrics was the consultation of a content expert (SH) who could act as a representative of the DHOH community with the intent of gauging the quality and applicability of DHOH apps. As a deaf individual, he was able to advocate for the needs of the target population from a position of membership; the study gained insight into the needs of the population along with reasonable knowledge to design content-specific measures included in the MARS. This proved invaluable in guiding the direction of the study to focus not only on the features of the apps that would be representative of a quality app but also on those that could augment the interpersonal and social capability of the user.

Strengths and Limitations

The study leverages the MARS, which is considered a valid and reliable scale for evaluating mHealth apps. The study introduces a novel MARS modification to create content-specific criteria using a content expert to address the needs of intended end users who will use the apps. One of the limitations of this review was that the search was limited to US app stores. This limitation might have restricted the results and the quality of the apps for the DHOH, particularly if other countries are further ahead in the development of DHOH apps. Only DHOH apps that are publicly available were included, which could have excluded apps developed by a specific health care network that focuses on the DHOH population. A relatively small sample of apps was included in the study. The entire pool of apps for the DHOH was not large to begin with, which may be representative of the DHOH population being a smaller segment of the entire population. There was a high rate of DHOH app turnover, which might be due to the information quality, or that the app is not meeting the needs of the DHOH population**.**

Future Research

The authors of this study suggest continuing research into the use of the modified MARS scale with content-specific (app-specific) criteria developed with content experts to evaluate mHealth apps for different populations. Further research into the use of content experts while designing a study, and their effect on the validity of subject-specific content will elucidate more information about the potential benefits of using content experts to design content-specific metrics. Additional research is needed for both mHealth apps and DHOH apps to establish additional criteria to measure information quality in terms of patient safety and privacy. Being able to measure the risk of patient safety is a growing concern as some mHealth apps do not follow evidence-based guidelines, or the app developers have little or no medical training [42,43]. Other research opportunities would be to design some app evaluation criteria that can aggregate different types of apps with some beneficial DHOH features (not directly intended for DHOH users). For example, in this study, the initial search terms identified apps such as audio analyzers and sound detectors, which were excluded as they did not have DHOH features. However, 1 sound detection app, eezySoundDetector, recognized the potential to assist the DHOH by providing flashing lights or vibrating alarms for a fire or a crying child. There were other types of apps that were excluded for only having minimal DHOH features, yet these might make a difference for the DHOH in a predominantly hearing world. Developing an evaluation tool that can aggregate a wider variety of apps with different assistive features could be a significant contribution to the development of mHealth apps. In addition, adding a qualitative component to studies that survey user preferences and feedback on the usability of specific mHealth apps has shown to be effective. A preliminary study of a sound detection algorithm that used a training and feedback survey from the DHOH got firsthand insights about detection preferences, such as running water in the home or a printer in the work environment, which improve app usability [44]. Finally, this study can be expanded with a follow-up review on the highest ranked DHOH apps to access effectiveness and tangible gains identified by the users, as well as to measure the DHOH app turnover.

Conclusions

The MARS remains a high-fidelity tool for app evaluation. The study emphasizes the value of including a content expert early in the mHealth app development process as well as the evaluation process to improve effectiveness and to assist in making criteria-based recommendations to end users. The focus on mHealth apps for the DHOH population illustrates the importance of including a related content expert. For someone who has a hearing difficulty, it can be both reassuring and empowering to see mHealth app developers, evaluators, and providers, who recommend these products, value the understanding of the needs of the intended users. This can be particularly true for an individual with hearing difficulties who struggles with his or her identity in a world set in an overwhelmingly hearing context.

Conflicts of Interest

None declared.

References

1. Sapci A, Sapci H. Digital continuous healthcare and disruptive medical technologies: m-Health and telemedicine skills training for data-driven healthcare. J Telemed Telecare 2018 Aug 22:1357633X18793293. [doi: 10.1177/1357633X18793293] [Medline: 30134779]
2. Terry M. Medical apps for smartphones. Telemed J E Health 2010;16(1):17-22. [doi: 10.1089/tmj.2010.9999] [Medline: 20070172]

3. Statista. Number of Mobile App Downloads Worldwide From 2016 to 2018 URL: https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/ [accessed 2019-02-23]
4. Census. Americans With Disabilities: 2010 URL: https://www.census.gov/library/publications/2012/demo/p70-131.html [accessed 2018-12-12]
5. Carroll DD, Courtney-Long EA, Stevens AC, Sloan ML, Lullo C, Visser SN, Centers for Disease Control and Prevention (CDC). Vital signs: disability and physical activity--United States, 2009-2012. MMWR Morb Mortal Wkly Rep 2014 May 9;63(18):407-413 [FREE Full text] [Medline: 24807240]
6. Americans with Disabilities Act. URL: https://www.ada.gov/ [accessed 2019-02-23]
7. Centers for Disease Control and Prevention. Disability and Health Inclusion Strategies URL: https://www.cdc.gov/ncbddd/disabilityandhealth/disability-strategies.html [accessed 2018-12-12]
8. Jaeger PT, Xie B. Developing online community accessibility guidelines for persons with disabilities and older adults. J Disabil Policy Stud 2008 Oct 17;20(1):55-63. [doi: 10.1177/1044207308325997]
9. Maalej W, Kurtanović Z, Nabil H, Stanik C. On the automatic classification of app reviews. Requir Eng 2016;21(3):311-331. [doi: 10.1007/s00766-016-0251-9]
10. 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]
11. Reagan T. Toward an 'archeology of deafness': Etic and Emic constructions of identity in conflict. J Lang Identity Educ 2002 Jan 2;1(1):41-66. [doi: 10.1207/S15327701JLIE0101_4]
12. Jones M. Deafness as culture: a psychosocial perspective. Disabil Stud Q 2002 Apr 15;22(2). [doi: 10.18061/dsq.v22i2.344]
13. National Association of the Deaf. Community and Culture – Frequently Asked Questions URL: https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/ [accessed 2019-07-06]
14. Morgan G. On language acquisition in speech and sign: development of combinatorial structure in both modalities. Front Psychol 2014;5:1217 [FREE Full text] [doi: 10.3389/fpsyg.2014.01217] [Medline: 25426085]
15. Kacorri H, Huenerfauth M, Ebling S, Patel K, Menzies K, Willard M. Regression analysis of demographic and technology-experience factors influencing acceptance of sign language animation. ACM Trans Access Comput 2017;10(1):1-33. [doi: 10.1145/3046787]
16. Mitchell RE, Young TA, Bachleda B, Karchmer MA. How many people use ASL in the United States? Why estimates need updating. Sign Lang Stud 2006;6(3):306-335. [doi: 10.1353/sls.2006.0019]
17. McKee MM, Winters PC, Sen A, Zazove P, Fiscella K. Emergency department utilization among deaf American sign language users. Disabil Health J 2015 Oct;8(4):573-578 [FREE Full text] [doi: 10.1016/j.dhjo.2015.05.004] [Medline: 26166160]
18. Bureau of Labor and Statistics. Interpreters and Translators-Job Outlook URL: https://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm#tab-6 [accessed 2019-02-23]
19. HandSpeak. How Long Does It Take to Learn ASL? URL: https://www.handspeak.com/learn/index.php?id=61 [accessed 2019-02-23]
20. National Association of the Deaf. URL: https://www.nad.org/ [accessed 2019-02-23]
21. Pang H. How does time spent on WeChat bolster subjective well-being through social integration and social capital? Telemat Inform 2018 Dec;35(8):2147-2156. [doi: 10.1016/j.tele.2018.07.015]
22. Chan M. Mobile phones and the good life: Examining the relationships among mobile use, social capital and subjective well-being. New Media Soc 2015;17(1):96-113. [doi: 10.1177/1461444813516836]
23. Assistive Technology Industry Association. URL: https://www.atia.org/ [accessed 2018-05-30]
24. Gaffney M, Eichwald J, Grosse S, Mason C. Identifying infants with hearing loss-United States, 1999-2007. Morb Mortal Wkly Rep 2010;59(8):220-223 [FREE Full text]
25. The National Institute on Deafness and Other Communication Disorders. 2015. American Sign Language URL: https://www.nidcd.nih.gov/health/american-sign-language [accessed 2019-02-23]
26. Dickson M, Magowan R, Magowan R. Meeting deaf patients' communication needs. Nurs Times 2014;110(49):12-15. [Medline: 26016132]
27. Gannon C. The deaf community and sexuality education. Sex Disabil 1998;16(4):283-293. [doi: 10.1023/A:1023067828478]
28. Ahmed T, Wahid M, Habib M. Implementation of Bangla Speech Recognition in Voice Input Speech Output (VISO) Calculator. In: Proceedings of the 2018 International Conference on Bangla Speech and Language Processing. 2018 Presented at: ICBSLP'18; September 21-22, 2018; Sylhet, Bangladesh. [doi: 10.1109/icbslp.2018.8554773]
29. Strong M, Prinz PM. A study of the relationship between American sign language and English literacy. J Deaf Stud Deaf Educ 1997;2(1):37-46. [doi: 10.1093/oxfordjournals.deafed.a014308] [Medline: 15579834]
30. YouTube. 2018. MARS training video URL: https://www.youtube.com/watch?v=25vBwJQIOcE [accessed 2018-04-26]
31. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
32. Krippendorff K. University of Pennsylvania ScholarlyCommons. 2011. Computing Krippendorff 's Alpha-Reliability URL: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1043&context=asc_papers [accessed 2019-10-01]

33. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
34. Liddell S. Grammar, Gesture, And Meaning In American Sign Language. Cambridge: Cambridge University Press; 2003.
35. Emmorey K, Damasio H, McCullough S, Grabowski T, Ponto LL, Hichwa RD, et al. Neural systems underlying spatial language in American Sign Language. Neuroimage 2002 Oct;17(2):812-824. [doi: 10.1006/nimg.2002.1187] [Medline: 12377156]
36. Sunjoto M, Unwin J. UCL Discovery. 2017. Traffic signage conspicuity URL: http://discovery.ucl.ac.uk/1570053/1/Unwin_LJ%20July_Aug%20p30-35%20new-ilovepdf-compressed.pdf [accessed 2019-10-01]
37. Pisharady PK, Vadakkepat P, Loh AP. Attention based detection and recognition of hand postures against complex backgrounds. Int J Comput Vis 2013;101(3):403-419. [doi: 10.1007/s11263-012-0560-5]
38. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
39. Sharpe M. CORE. 2010. Assistive Technology Attrition: Identifying Why Teachers Abandon Assistive Technologies URL: https://core.ac.uk/download/pdf/51097930.pdf [accessed 2019-10-01]
40. Verza R, Carvalho ML, Battaglia MA, Uccelli MM. An interdisciplinary approach to evaluating the need for assistive technology reduces equipment abandonment. Mult Scler 2006 Feb;12(1):88-93. [doi: 10.1191/1352458506ms1233oa] [Medline: 16459724]
41. Kamau PW, Ivey SL, Griese SE, Qari SH. Preparedness training programs for working with deaf and hard of hearing communities and older adults: lessons learned from key informants and literature assessments. Disaster Med Public Health Prep 2018 Oct;12(5):606-614. [doi: 10.1017/dmp.2017.117] [Medline: 29041996]
42. Ferrero NA, Morrell DS, Burkhart CN. Skin scan: a demonstration of the need for FDA regulation of medical apps on iPhone. J Am Acad Dermatol 2013 Mar;68(3):515-516. [doi: 10.1016/j.jaad.2012.10.045] [Medline: 23394920]
43. Huckvale K, Car M, Morrison C, Car J. Apps for asthma self-management: a systematic assessment of content and tools. BMC Med 2012 Nov 22;10:144 [FREE Full text] [doi: 10.1186/1741-7015-10-144] [Medline: 23171675]
44. Bragg D, Huynh N, Ladner R. A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users. In: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility. 2016 Presented at: ASSETS'16; October 23-26, 2016; Reno, Nevada, USA p. 3-13. [doi: 10.1145/2982142.2982171]

Abbreviations

ASL: American Sign Language
AT: assistive technology
DHOH: deaf and hard-of-hearing
GUI: graphical user interface
MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by G Eysenbach; submitted 29.03.19; peer-reviewed by R Wolfe, R Dewey; comments to author 15.06.19; revised version received 25.07.19; accepted 18.08.19; published 30.10.19.

Please cite as:
Romero RL, Kates F, Hart M, Ojeda A, Meirom I, Hardy S
Quality of Deaf and Hard-of-Hearing Mobile Apps: Evaluation Using the Mobile App Rating Scale (MARS) With Additional Criteria From a Content Expert
JMIR Mhealth Uhealth 2019;7(10):e14198
URL: http://mhealth.jmir.org/2019/10/e14198/
doi:10.2196/14198
PMID:31670695

Review

Electronic Health Interventions to Improve Adherence to Antiretroviral Therapy in People Living With HIV: Systematic Review and Meta-Analysis

Ziqi Wang[1], MD; Yaxin Zhu[1], MD; Liyuan Cui[2], MD; Bo Qu[1], MD, PhD

[1]School of Public Health, China Medical University, Shenyang, China

[2]School of Medical Informatics, China Medical University, Shenyang, China

Corresponding Author:
Bo Qu, MD, PhD
School of Public Health
China Medical University
77 Puhe Road
Shenyang North New Area
Shenyang, 110122
China
Phone: 86 18900910198
Email: bqu@cmu.edu.cn

Abstract

Background: Electronic health (eHealth) is increasingly used for self-management and service delivery of HIV-related diseases. With the publication of studies increasingly focusing on antiretroviral therapy (ART) adherence, this makes it possible to quantitatively and systematically assess the effectiveness and feasibility of eHealth interventions.

Objective: The purpose of this review was to explore the effectiveness of eHealth interventions on improving ART adherence in people living with HIV. The effects of different intervention characteristics, participant characteristics, and study characteristics were also assessed.

Methods: We systematically searched MEDLINE (via PubMed), EMBASE, the Cochrane Central Register of Controlled Trials, and 3 conference abstract databases using search terms related to HIV, ART, adherence, and eHealth interventions. We independently screened the studies, extracted the data, and assessed the study quality and then compared the information in pairs. Articles published in English that used randomized controlled trials to assess eHealth interventions to improve ART adherence of people living with HIV were identified. We extracted the data including study characteristics, participant characteristics, intervention characteristics, and outcome measures. The Cochrane risk-of-bias tool was used to assess the risk of bias and study overall quality. Odds ratios, Cohen *d*, and their 95% CIs were estimated using random-effects models. We also performed multiple subgroup analyses and sensitivity analyses to define any sources of heterogeneity.

Results: Among 3941 articles identified, a total of 19 studies (including 21 trials) met the inclusion criteria. We found 8 trials from high-income countries and 13 trials from low- and middle-income countries. Furthermore, at baseline, the health status of participants in 14 trials was healthy. Of the trials included, 7 of 21 used personality content, 12 of 21 used a 2-way communication strategy, and 7 of 21 used medical content. In the pooled analysis of 3937 participants (mean age: 35 years; 47.16%, 1857/3937 females), eHealth interventions significantly improved the ART adherence of people living with HIV (pooled Cohen *d*=0.25; 95% CI 0.05 to 0.46; *P*=.01). The interventions were also correlated with improved biochemical outcomes reported by 11 trials (pooled Cohen *d*=0.25; 95% CI 0.11 to 0.38; *P*<.001). The effect was sensitive to sample size (*Q*=5.56; *P*=.02) and study duration (*Q*=8.89; *P*=.003), but it could not be explained by other moderators. The primary meta-analysis result was stable in the 3 sensitivity analyses.

Conclusions: Some of the eHealth interventions may be used as an effective method to increase the ART adherence of people living with HIV. Considering that most of the trials included a small sample size and were conducted for a short duration, these results should be interpreted with caution. Future studies need to determine the features of eHealth interventions to better improve ART adherence along with long-term effectiveness of interventions, effectiveness of real-time adherence monitoring, enhancement of study design, and influences on biochemical outcomes.

(JMIR Mhealth Uhealth 2019;7(10):e14404) doi:10.2196/14404

KEYWORDS

HIV; highly active antiretroviral therapy; medication adherence; eHealth

Introduction

Background

Owing to the significant role of antiretroviral therapy (ART) in controlling HIV from 2000 to 2017, AIDS-related deaths decreased by 38%, and approximately 11.4 million lives were saved [1]. Although ART has achieved great success, the 2030 treatment targets of the new 90-90-90 of the Joint United Nations Programme on HIV/AIDS appear unachievable for many countries [2]. In 2017, 21.7 million people living with HIV (PLWH) received ART, which accounted for only 59% of the global PLWH and 52% of children living with HIV [3]. In addition, according to the data from World Health Organization (WHO) in 2016, less than 50% of the PLWH achieved viral suppression [4]. Patients who strictly adhere to ART could control disease progression and prevent the emergence of drug-resistant mutations [5]. Poor ART adherence lead to the accelerated progression of PLWH to AIDS [6,7], increased demand for medical interventions [8], increased morbidity and mortality [9], and increased circulating ART-resistant strains [7,10]. A number of traditional measures can be used to ensure ART adherence, including behavioral skills training or medication adherence training, cognitive behavioral therapy, peer or social support, and counseling [11]. However, most of the interventions that are used in long-term therapy are either complicated or not widely applicable, and thus, more convenient, low-cost, and widely feasible innovations are required [8,12].

Owing to advances in mobile phone and internet technologies, the use of electronic health (eHealth) is expanding. The WHO Global Observatory for eHealth defines eHealth as “the use of information and communication technologies (ICT) for health” [13]. This involves the delivery of health information for health professionals and consumers through telecommunication (short message service, SMS; patient monitoring devices; and mobile phones) and internet-based components (social media, computer software, websites, mobile apps, games, and chat rooms) [14]. Numerous barriers to PLWH remain, including persistent stigma and discrimination [15], low socioeconomic status [16], punitive laws [17], and geographical isolation [18]. eHealth is increasingly used for the self-management and service delivery of HIV-related diseases [19]. eHealth interventions have many advantages: eHealth interventions are low cost and suitable for use in low- and middle-income countries (LMICs) [19-21] as well as convenient and accessible. According to estimates by the Ericsson 2018 Mobility Report, the number of mobile subscriptions worldwide will reach 7.9 billion in the third quarter of 2018 [22]. Moreover, popular social media platforms including WeChat [23], Facebook, and YouTube [24,25] have more than 1 billion monthly active users. In addition, eHealth can provide users with a private space to remove the discrimination and stigma associated with HIV [26,27]. eHealth can also boost behavioral changes, self-efficacy, knowledge, and clinical outcomes and has been developed for a wide range of disease and health behaviors [28-30].

In view of these advantages, an increasing number of reviews have studied the effects of eHealth on the promotion of ART adherence of PLWH. Therefore, in this study, before we conducted formal systematic literature search, a literature search was performed in MEDLINE to identify systematic reviews and meta-analyses published before March 20, 2018, that reviewed eHealth interventions to improve ART adherence (search terms are shown in Multimedia Appendix 1). Although favorable effects of eHealth interventions were documented, only narrative and systematic reviews were reported [31-34]. Moreover, additional reviews were either targeted to only 1 type of eHealth (such as SMS [7,35-37], social media [26], and voice calls [37]) or were only performed in the specific group of participants (men who have sex with men [38] and key populations in the Asia-Pacific region [19]).

Objectives

With the publication of more and more studies focusing on ART adherence, this makes it possible to make quantitative and systematic assessments of the effectiveness and feasibility of eHealth interventions. In addition, despite the diversity of the interventions, we aggregated and compared their effects on improving ART adherence, which was supported by functional similarity and characteristics. So, the primary purpose of this study was to explore the effectiveness of eHealth interventions on improving ART adherence of PLWH. Moreover, the effects of different intervention characteristics, participant characteristics, and study characteristics were also assessed. To enhance the methodological quality of the meta-analysis and strengthen the conclusions, only randomized controlled trials (RCTs) were included.

Methods

Guidelines

This review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [39] and the Cochrane Collaboration reporting items for systematic reviews and meta-analyses [40].

Literature Search

We systematically searched MEDLINE (via PubMed), EMBASE, and the Cochrane Central Register of Controlled Trials for relevant studies published in English without restriction on publication date. The date of the last search for the electronic database was March 25, 2018. At the same time, we also searched for abstracts on several conference databases including the International AIDS Conference, the International AIDS Society Conference on HIV Science, and the Conference on Retroviruses and Opportunistic Infections. The reference lists of all relevant studies were searched manually to identify potential trials. The search strategy was developed by a librarian

(LC) to identify studies that used eHealth interventions to improve ART adherence of PLWH. The study was developed based on Medical Subject Headings and key terms related to 4 categories: HIV, ART, adherence, and eHealth interventions. Detailed search items are listed in Multimedia Appendix 2.

Selection Criteria

By following the populations, interventions, comparisons, outcomes, and study design (PICOS) framework, we included trials when (1) the study population was targeted to a sample of PLWH on ART; (2) the intervention focused only on eHealth interventions aimed to increase ART adherence rather than a data collection or participant recruitment tool; (3) the control group was the usual standard of care for PLWH; (4) the outcomes reported at least one ART adherence measurement (self-report, pill counting, electronic drug monitoring devices, or pharmacy refill record) and biochemical outcomes (viral load, $\log_{10}$ copies/mL, cluster of differentiation 4^+ cell ($CD4^+$) counting, or viral suppression [VS]/ virological failure [VF]); and (5) the study design was an RCT with a minimum of 3 months follow-up. No restrictions on the treatment of the participants, previous ART failure, or geography were applied. If multiple studies were reported on the same trial, the study with the most relevant outcome was included. Detailed PICOS criteria for the included studies are listed in Multimedia Appendix 3.

Data Extraction

A total of 2 authors (ZW and YZ) independently reviewed all the titles and abstracts of the initial literature using bibliographic citation management software (EndNote, Version X7, Thomson Reuters) to determine their relevance based on the above-mentioned selection criteria. Relevant studies were kept for full-text reviews. Discrepancies were resolved by discussion with a third independent researcher (BQ).

Using a standardized extraction form (Microsoft Office Excel, Version 2013), the same 2 authors independently performed data extraction based on the following information: study characteristics (first author or research team, research year, setting, location, and study duration); participant characteristics (sample size [intervention arm, IA/control arm, CA], mean age, female ratio, and participant inclusion criteria); intervention characteristics (intervention type [IA/CA], frequency of intervention, intervention content [general content/medical content], personalization, and intervention communication strategy [1-way/2-way]); and outcome measures (primary adherence outcome measure [the proportion of medication taken as prescribed/the proportion of adherent patients], adherence outcome assessment methods, and biochemical outcome assessment methods). For studies with multiple IAs, eligible comparison trials were extracted and divided into distinct trials based on recent guidelines [40]. When a multiple-phase follow-up was reported, the outcome of the final follow-up corresponding to the study duration was used to assess the persistence and sustainability of the intervention [6,7]. The data of outcome measures were used to calculate the effect size in the meta-analysis. If there was insufficient data to calculate the effect size, the corresponding author was contacted by email. If the data were unavailable, studies were excluded [41]. For studies reporting a median and interquartile range for adherence outcomes, we converted the outcomes into the mean (SD) as previously reported [42,43].

Assessment of Study Quality

Methodological quality is an important facet of this review. ZW and YZ independently assessed the risk of bias within individual included studies using the Cochrane risk-of-bias tool [40], which recommends 7 dimensions of research methodology for RCTs: (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of the outcome assessment, (5) incomplete outcome data, (6) selective outcome reporting, and (7) other sources of bias. The risk of bias for each item was evaluated at 3 levels: (1) high, (2) unclear, or (3) low. If a study was evaluated as a high or unclear risk of bias for sequence generation or randomization concealment, and other dimensions had more than 2 high risks of bias, the studies were considered as low overall quality. A third author (BQ) collated the results. Detailed quality assessments for the included studies are listed in Multimedia Appendix 4.

Statistical Analysis

Statistical Methods

Statistical analyses of this meta-analysis were performed using the CMA Software (Comprehensive Meta-analysis, Version 2, Biostat). We used the mean effect size approach to pool estimates, which have been applied in other studies [7,8]. The effect size was weighted as per the study sample size. We calculated the odds ratio (OR) and the 95% CI for each included trial. Random-effects models were used to pool estimates as large between-study heterogeneity was expected. Cohen *d* values and the 95% CI were used to calculate the magnitude of the effect size. Values of 0.2, 0.5, and 0.8 were considered small, medium, and large effect sizes, respectively [8]. Reported *P* values were 2-tailed. To assess heterogeneity, I^2 and Q statistics were used. I^2 statistic exceeding 50% with a significant *Q* value ($P<.05$) represented substantial heterogeneity [44]. I^2 also represented the levels of heterogeneity with values of 25%, 50%, and 75% indicating low, moderate, and high heterogeneity, respectively [45]. Funnel plot symmetry [46] and Egger regression intercept [47] were used to assess publication bias. If publication bias existed, the funnel plots were asymmetric (Egger test: $P<.05$). We used trim-and-fill analysis described by Duval and Tweedie to estimate the number of missing studies because of publication bias and calculated the effect size after correction [48].

Weighted mean effect sizes were calculated to estimate the overall difference between eHealth and control groups on adherence outcomes as well as on biochemical outcomes because biochemical outcomes are the final presentation of adherence. Of the trials that reported different outcomes, the majority of trials (14/21) reported multiple adherence outcomes, and more than half of the trials (6/11) reported multiple biochemical outcomes. Considering that multiple effect sizes in 1 trial violated the independence assumption in meta-analysis, we selected only 1 effect size for each trial in our analyses.

When trials had multiple outcome assessment methods, we selected the most objective and reliable method according to a predetermined order (assessed in the following order: electronic monitoring, pill counting, pharmacy refill record, self-report, and treatment interruption; viral load, $\log_{10}$ copies/mL, $CD4^+$ cell counting, and VS/VF; and continuous scale over dichotomized scale) as used in other studies [8,49]. The mean effect size was also independently calculated in all the adherence assessment methods and the biochemical outcome assessment methods of the trials.

Subgroup Analyses

Given the potential for substantial significant heterogeneity across the studies (based on the I^2 and Q statistics for heterogeneity), we performed subgroup analyses to explore the potential factors that moderate the overall effect size. The following moderators were examined: age (age <36.65 years or age ≥36.65 years), study duration (short-term trial: duration ≤36 weeks or long-term trial: duration >36 weeks), sample size (large trial: n≥166 or small trial: n<166), location (high-income countries or LMICs), participant ART status at baseline (nonadherence, ART-naïve, or treatment experienced), participant health status at baseline (healthy or at risk), age category (adults or adults and adolescents), primary outcome measure (proportion of medication taken as prescribed or proportion of patients with good adherence), type of intervention (Web-based computer programs, telephone calls, SMS, electronic adherence monitoring device [EAMD], or SMS plus telephone calls), frequency of intervention (real-time, daily, or frequency below daily), intervention content (medical content or general content), communication strategy (1-way or 2-way), and personalization (yes or no). In particular, we also divided the type of intervention into telecommunication subgroup and internet-based component subgroup so that we could explore whether there was a notable and significant difference between the 2 subgroups. The cut-off points for moderators (age, sample size, and study duration) were based on the median values among trials from the available information, which was used by several previous studies [8,49].

Sensitivity Analyses

A total of 3 sensitivity analyses were performed to assess stability of the meta-analysis. The first sensitivity analysis excluded low-quality trials, the second excluded trials with a sample attrition rate ≥20%, and the third gave higher weight to specific assessment methods (self-report and $CD4^+$ cell counting) for the trials reporting multiple outcome measures.

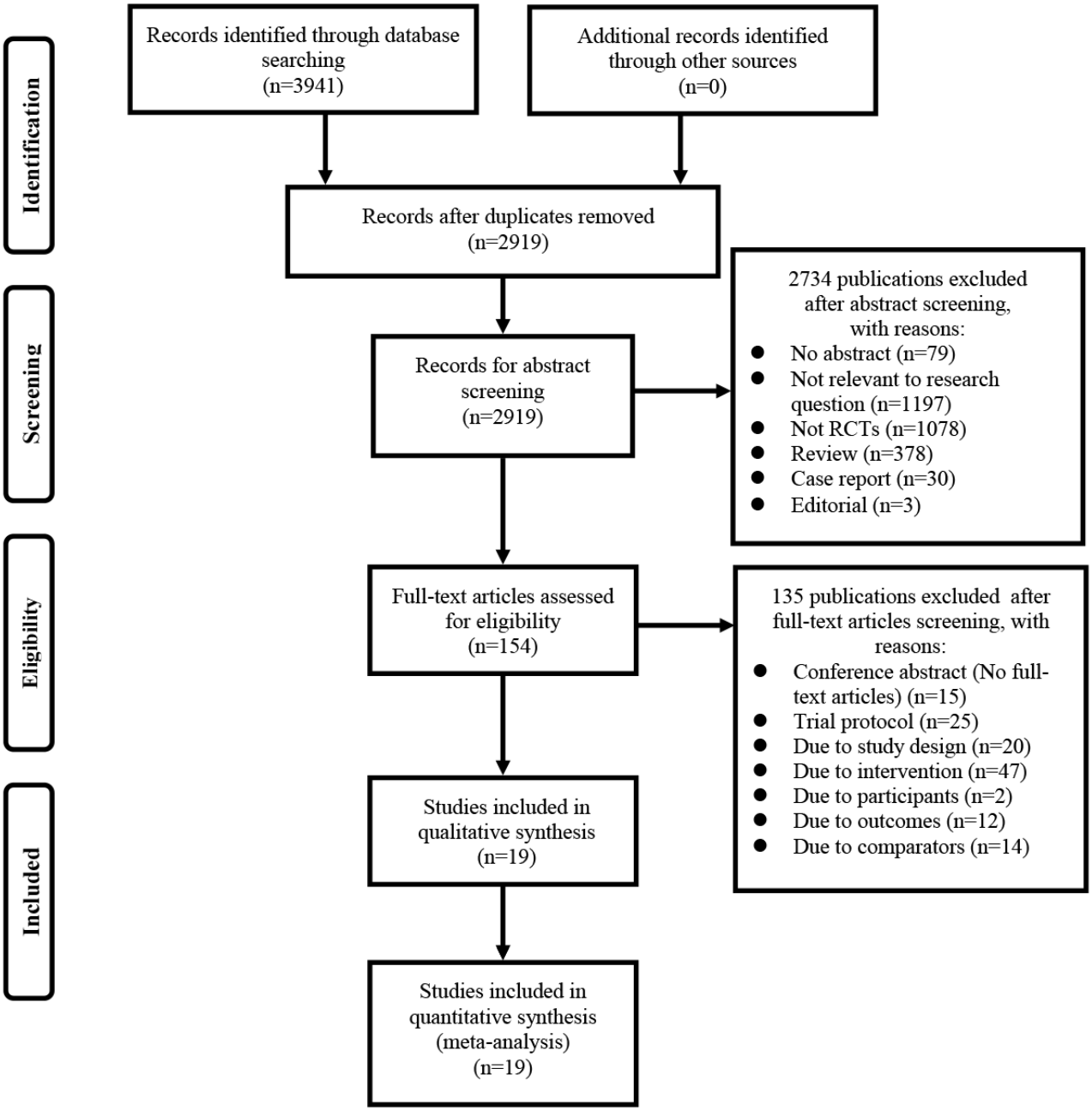
Results

Study Characteristics

A total of 19 RCTs were identified following the assessment of 154 full-text articles (Figure 1; Multimedia Appendices 5-8) [50-68]. We extracted 2 independent comparison trials (daily SMS and weekly SMS) from the study of Pop-Eleches et al [54] and 2 comparison trials (1-way and 2-way communication strategies) from the study of Linnemayr et al [68]. Finally, a total of 21 trials were included in the meta-analysis consequently. A total of 21 trials included 3937 participants. The sample size varied from 21 to 631, with a median of 166. The mean age of the participants was 35 years (Safren et al failed to report the mean age of participants [50]), and 47.16% (1857/3937) were female. Studies were performed in the United States [50-52,57,59,61,62,64], Kenya [53,54], China [60,65], Uganda [68], Brazil [56], India [58], Cameroon [55], South Africa [63], Botswana [66], and Malaysia [67]. Study duration ranged from 12 to 96 weeks, with a median of 36 weeks. One-third of the trials targeted at-risk populations (7/21), 43% focused on ART-naïve populations (9/21), 76% focused on adults (16/21), and the remainder focused on adults and adolescents (5/21).

The purpose of the included studies was to improve ART adherence of PLWH. Self-report (10/21) and electronic drug monitoring device (medication event monitoring system cap and EAMD [Wisepill]; 10/21) were the most commonly used methods to assess adherence, followed by pill counting (2/21), pharmacy refill record (2/21), or treatment interruption (4/21). Primary type of outcome measure was presented as the proportion of medication taken as prescribed in 15 trails and as the proportion of patients with good adherence in 6 trials. Biochemical outcomes were measured through $CD4^+$ cell counting (6/21), viral load, $\log_{10}$ copies/mL (5/21), and VS/VF (6/21).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses chart showing study selection process. RCT: randomized controlled trial.



Electronic Health Characteristics

The eHealth characteristics are varied across the 21 trials. A total of 12 trials sent SMSs, 4 used telephone calls, 2 performed interventions in Web-based computer program, 2 used EAMD, and 1 combined SMS with telephone calls. According to another classification method of intervention, 19 trials were divided into telecommunication subgroup, and the remaining 2 trials were divided into internet-based component subgroup. A total of 19 trials conducted interventions at a fixed predetermined frequency (daily or frequency below daily), and the remaining 2 trials used real-time medication monitoring in which the participants were sent a reminder if they did not open the medication management device within the specified time. The intervention content was general content in 14 trials (medication reminders, humor jokes, as well as motivation and encouragement) and medical content in 7 trials (HIV/AIDS-related knowledge, the importance of adherence, and skills for good adherence). Moreover, 12 trials supported a 2-way communication strategy (patients were permitted, encouraged, or demanded to respond to the received information). Furthermore, 7 trials used personalized content (eg, the trial by Simoni et al [52] used the flexible content of messages to accommodate the different needs and schedules of the participants).

Meta-Analyses

Table 1 shows the mean effect sizes across all types of outcome assessment methods. Statistical significance for the individual outcome assessment method was not always achieved because of the limited statistical power of the available studies. For the 5 adherence outcome assessment methods, significant results

in both self-report (k=10; Cohen d=0.44; 95% CI 0.11 to 0.77; P=.01) and pharmacy refill record (k=2; Cohen d=0.47; 95% CI 0.11 to 0.84; P=.01) were observed. For the 3 biochemical outcome assessment methods, $CD4^+$ cell counting had a small positive effect size (k=6; Cohen d=0.20; 95% CI 0.04 to 0.35; P=.01), and viral load (log_{10} copies/mL) had a negative significant effect size (k=5; Cohen d=−0.40; 95% CI −0.62 to −0.17; P<.001).

Table 1. The effect of electronic health on antiretroviral therapy adherence outcomes and biochemical outcomes by type of outcome assessing methods.

Measures	k (number of trials)	Odds ratio (95% CI)	Cohen d (95% CI)	P value	I^2 (%)
Electronic drug monitoring device	10	1.20 (0.75 to 1.93)	0.10 (−0.16 to 0.36)	.46	80.44
Self-report	10	2.20 (1.21 to 4.00)	0.44 (0.11 to 0.77)	.01	88.52
Pill counting	2	0.79 (0.52 to 1.21)	−0.13 (−0.36 to 0.10)	.28	2.82
Pharmacy refill record	2	2.36 (1.22 to 4.56)	0.47 (0.11 to 0.84)	.01	0.00
Treatment interruption	4	0.69 (0.41 to 1.15)	−0.21 (−0.49 to 0.08)	.15	0.00
Cluster of differentiation 4^+ cell counting	6	1.43 (1.08 to 1.89)	0.20 (0.04 to 0.35)	.01	21.94
Viral load (log_{10} copies/mL)	5	0.49 (0.32 to 0.73)	−0.40 (−0.62 to −0.17)	<.001	30.83
Viral suppression/virological failure	6	1.32 (0.90 to 1.93)	0.15 (−0.06 to 0.36)	.16	34.53
Mean biochemical outcomes	11	1.57 (1.22 to 2.01)	0.25 (0.11 to 0.38)	<.001	43.16
Mean adherence outcomes	21	1.59 (1.10 to 2.29)	0.25 (0.05 to 0.46)	.01	86.70

In the pooled analysis of 21 trials, eHealth interventions significantly improved ART adherence (OR=1.59, 95% CI 1.10 to 2.29; P=.01; Figure 2). The weighted mean effect size (Cohen d) was 0.25 (95% CI 0.05 to 0.46). A small positive effect of eHealth interventions on improving ART adherence of PLWH was observed. Heterogeneity assessments showed variability across the trials (Q_{20}=150.36; P<.001). There was high heterogeneity (I^2: 86.70%) across trials, which supported the selection of the random-effects model to perform subgroup analyses to investigate the impact of the moderators on the overall effect size. Publication bias was not detected through funnel plot analysis (Figure 3) and Egger regression tests (Intercept of the regression line: 2.39; 95% CI –1.12 to 5.91; t_{19}=1.43; P=.17). Duval and Tweedie's trim-and-fill analysis showed that no studies were trimmed or filled, indicating no evidence of publication bias. In addition, biochemical outcomes reported by 11 trials also had a weighted mean effect size that achieved statistical significance (Cohen d=0.25; 95% CI 0.11 to 0.38; P<.001; Table 1).

Figure 2. The effect of electronic health intervention on antiretroviral therapy adherence of people living with HIV. Two independent comparison trials (daily short message service [SMS] and weekly SMS) from the study of Pop-Eleches et al were extracted as Pop-Eleches et al (1) and Pop-Eleches et al (2), and two independent comparison trials (1-way and 2-way communication strategies) from the study of Linnemayr et al were extracted as Linnemayr et al (1) and Linnemayr et al (2).

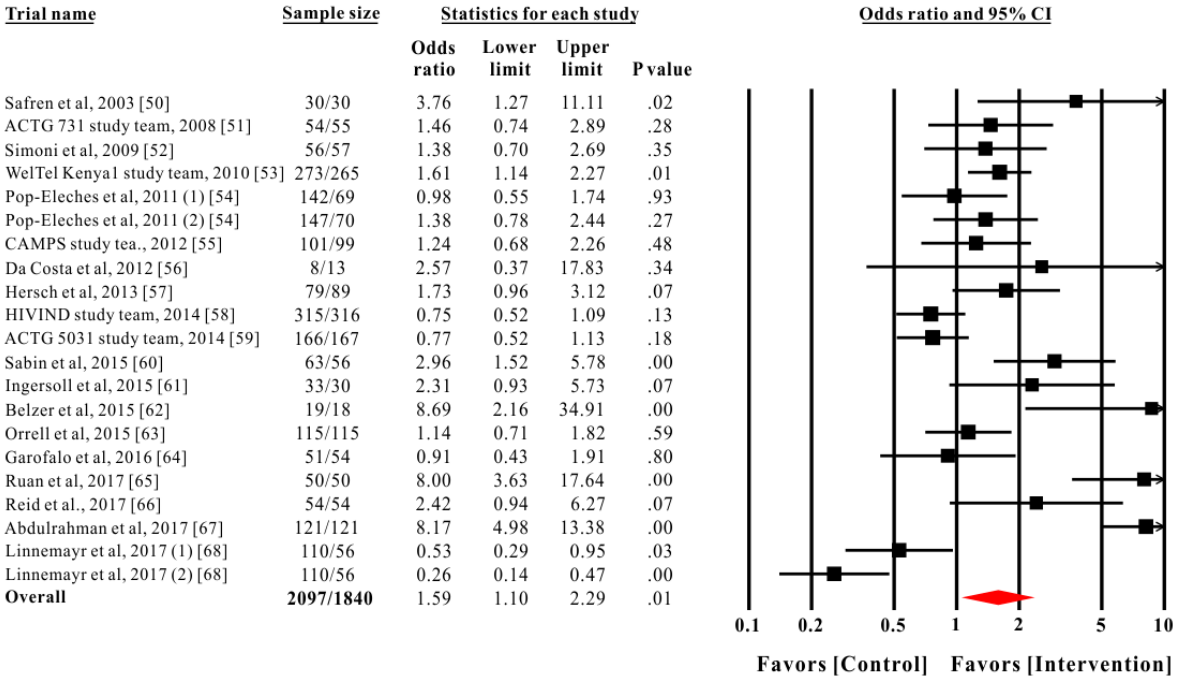
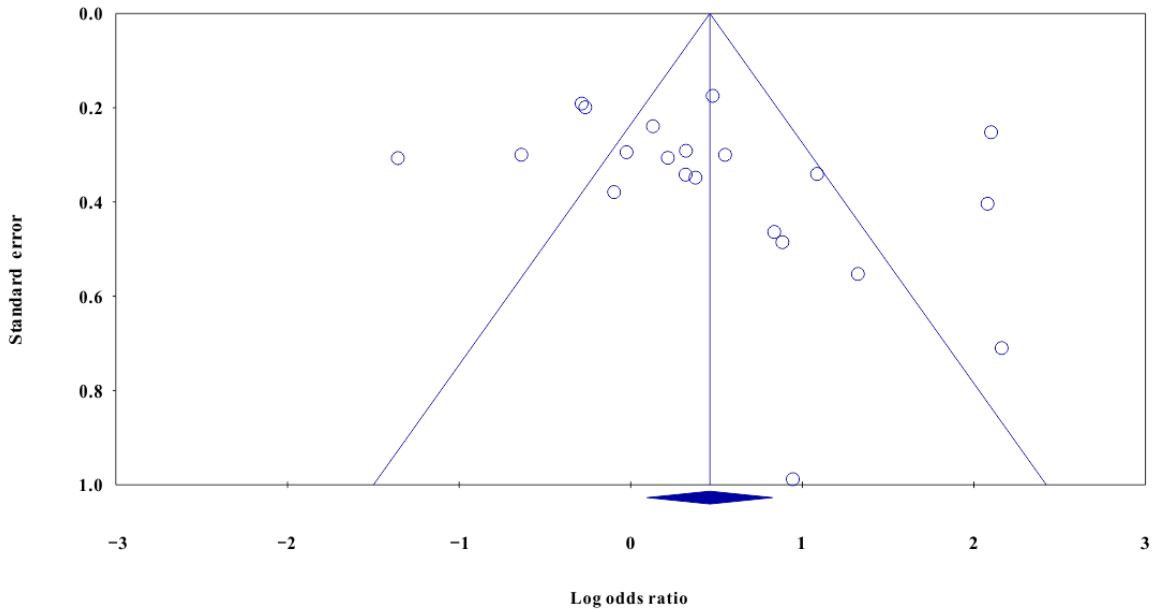


Figure 3. Funnel plot of SE and log odds ratio on antiretroviral therapy adherence of people living with HIV between intervention and control groups.



Subgroup Analyses

The study and participant characteristics could explain some heterogeneity across the trials, specifically the sample size and study duration (Table 2). The subgroup analysis of sample size showed that large trials (Cohen *d*=0.06; 95% CI −0.20 to 0.33) had smaller effect sizes than small trials (Cohen *d*=0.51; 95% CI 0.25 to 0.76), which showed a significant difference in ART adherence between these 2 subgroups (*Q*=5.56; *P*=.02). Short-term trials displayed medium effect sizes (Cohen *d*=0.51; 95% CI 0.23 to 0.79), whereas long-term trials showed no significant effect size (Cohen *d*=−0.01; 95% CI −0.22 to 0.19), which indicated a significant difference in ART adherence between the 2 subgroups (*Q*=8.89; *P*=.003). However, heterogeneity cannot be explained by mean age, location, participant ART status at baseline, participant health status at baseline, age category, primary type of outcome measure, and all the eHealth interventions characteristics (Tables 2 and 3).

Table 2. Subgroup analyses of the effect of electronic health on antiretroviral therapy adherence by study and participant characteristics.

Moderator and subgroups	*k* (number of trials)	Odds ratio (95% CI)	*Q* value	*P* value for heterogeneity
Sample size			5.59	.02
Large trial	11	1.12 (0.70 to 1.81)		
Small trial	10	2.50 (1.58 to 3.97)		
Mean age (years)			0.58	.45[a]
<36.65	9	1.28 (0.61 to 2.70)		
≥36.65	11	1.76 (1.25 to 2.49)		
Not specified	1	3.76 (1.23 to 11.48)		
Study duration			8.89	.003
Short-term trial	11	2.52 (1.53 to 4.16)		
Long-term trial	10	0.98 (0.67 to 1.42)		
Location			0.03	.86
High-income countries	8	1.62 (1.04 to 2.53)		
Low- and middle-income countries	13	1.52 (0.91 to 2.55)		
Participant ART[b] status at baseline			4.48	.11
ART-naïve	9	1.68 (0.98 to 2.89)		
Nonadherence	6	2.55 (1.39 to 4.66)		
Treatment experienced	6	0.97 (0.50 to 1.88)		
Participant health status at baseline			2.07	.15
At risk	7	2.24 (1.34 to 3.73)		
Healthy	14	1.36 (0.87 to 2.12)		
Age category			2.73	.10
Adults	16	1.89 (1.29 to 2.78)		
Adults and adolescents	5	0.89 (0.40 to 1.99)		
Primary type of outcome measure			0.92	.34
Proportion of medication taken as prescribed	15	1.71 (1.03 to 2.86)		
Proportion of patients with good adherence	6	1.26 (0.87 to 1.83)		

[a]The trial by Safren et al [50] did not report the mean age of participants.

[b]ART: antiretroviral therapy.

Table 3. Subgroup analyses of the effect of electronic health on antiretroviral therapy adherence by intervention characteristics.

Moderator and subgroup	*k* (number of trials)	Odds ratio (95% CI)	*Q* value	*P* value for heterogeneity
Type of intervention[a]			2.12	.55[b]
Web-based computer program	2	2.22 (1.09 to 4.51)		
Telephone call	4	1.21 (0.66 to 2.22)		
SMS[c]	12	1.31 (0.83 to 2.06)		
Electronic adherence monitoring device	2	1.78 (0.70 to 4.54)		
SMS plus telephone call	1	8.17 (4.98 to 13.38)		
Type of intervention[d]			0.83	.36
Telecommunication	19	1.53 (1.03 to 2.25)		
Internet-based component	2	2.22 (1.09 to 4.51)		
Frequency of intervention			0.29	.86
Real-time	2	1.78 (0.70 to 4.54)		
Daily	7	1.72 (1.10 to 2.70)		
Frequency below daily	12	1.44 (0.83 to 2.50)		
Intervention content			0.19	.67
General content	14	1.50 (0.93 to 2.42)		
Medical content	7	1.77 (0.97 to 3.23)		
Communication strategy			0.67	.41
1-way	9	1.91 (1.00 to 3.64)		
2-way	12	1.38 (0.90 to 2.13)		
Personalization			0.89	.34
Yes	7	1.31 (0.93 to 1.84)		
No	14	1.78 (1.04 to 3.07)		

[a]The intervention was divided into 5 subgroups: Web-based computer program, telephone call, short message service (SMS), electronic adherence monitoring device, and SMS plus telephone call.

[b]The trial by Abdulrahman et al [67] was the only one that used SMS plus telephone calls.

[c]SMS: short message service.

[d]The intervention was divided into 2 subgroups: telecommunication and internet-based component.

Sensitivity Analyses

The primary meta-analysis result was stable in the 3 sensitivity analyses. The effect of eHealth on improving ART adherence of PLWH did not change when we excluded 4 trials [50,51,61,63] with low quality (Cohen *d*=0.24; 95% CI 0.00 to 0.48; *P*=.049), when we excluded a trial [50] with sample attrition rates ≥20% (Cohen *d*=0.24; 95% CI 0.03 to 0.44; *P*=.02), and when we gave higher weight to self-report (replace the adherence assessment method of 2 trials—Simoni et al [52] and da Costa et al [56]—with the self-report; Cohen *d*=0.26; 95% CI 0.05 to 0.46; *P*=.01). In addition, the effect of eHealth on the biochemical outcomes of PLWH did not change when we gave higher weight to $CD4^+$ cell counting (replace the viral load, $\log_{10}$ copies/mL, of 3 trials—Simoni et al [52], Reid et al [66], and Abdulrahman et al [67]—with the $CD4^+$ cell counting; Cohen *d*=0.20; 95% CI 0.08 to 0.32; *P*<.001).

Discussion

Principal Findings

This review identified 21 trials of 19 RCTs that investigated the effectiveness of eHealth interventions on improving ART adherence of PLWH. Overall, eHealth interventions reported significant, but small, positive effects on ART adherence (Cohen *d*=0.25; 95% CI 0.05 to 0.46; *P*=.01) compared with PLWH in usual care. This finding was also stable in 3 sensitivity analyses. Specifically, SMS, telephone call, and EAMD were not able to significantly increase ART adherence of PLWH; however, the study of combining SMS and telephone call was highly effective in improving ART adherence. In addition, the Web-based computer program also showed significant positive effects in ART adherence. Our review also found that both telecommunication and internet-based components reported significant positive effects on ART adherence of PLWH. This meta-analysis result demonstrated that some of eHealth interventions showed favorable effects to improve ART

adherence of PLWH, which was consistent with findings of previous reviews. Daher et al [14] found that the digital innovations (mobile health, mHealth; internet-based mHealth/eHealth; and combined innovations) reported strong positive effects on improving ART adherence and clinic attendance rates. However, although significant, the small effect detected in this review was not sufficient to improve ART adherence and make it to the satisfactory clinical standard. Conn et al [69] mentioned in their study that the patient's medication adherence is difficult to change. The reasons for the ART nonadherence of PLWH are very complicated. In addition to the reason for forgetting to take medicine [70], it may also include psychological factors found in early reviews, such as depressive symptoms [70], stigma [71], and lack of social support [15]. Previous studies suggested that patients with chronic diseases may develop negative emotions during long-term medication and believe that their illnesses are incurable so that lacks the motivation to adhere to medication [12]. Moreover, nonadherence is also related to many factors including medication burden [72], side effects [73], and socioeconomic status [74]. Future research could try to use eHealth, educational, and psychosocial interventions together to better improve the ART adherence of PLWH.

Moderators on the Use of Electronic Health Interventions

Although this meta-analysis did not have significant publication bias, we noted significant heterogeneity, which may be because of clinical heterogeneity (the real difference of the impact generated from the different eHealth interventions and participant populations) or methodological heterogeneity (the difference generated from the different outcome assessment methods defined and measured in each study) [8]. Previous reviews on ART adherence also reported high heterogeneity [41]. Subgroup analyses showed that the effectiveness of eHealth interventions was sensitive to sample size and study duration. We found that small trials with limited sample sizes reported larger beneficial effects than large trials, which could be explained by the *small-study effects* proposed by Sterne et al [75]. Previous studies found that intervention effects were exaggerated in small trials with inadequate or unclear sequence generation, inadequate or unclear allocation concealment, and lack of blinding [76,77]. This is consistent with our findings that most of the trials with unclear sequence generation were small trials (4/5), and most of the small trials had unclear or high risk with blinding (9/10). The results of these small trials might overestimate the true effect of the interventions, and this effect is more easily published. Therefore, we should explain the results of the small trials with caution. Our subgroup analyses also indicated a higher effect size for short-term trials compared with long-term trials. This suggested that the effects of the eHealth interventions weakened over time. This finding is consistent with the findings of the study by Vervloet et al [78] who suggested that electronic reminders led to short-term improvements of the patients' adherence to medication, but the long-term effects were unclear. This finding has important clinical significance because the long-term effectiveness of eHealth interventions is a recent focus of attention. In addition, for the trials included in this review, most of them (19/21) were eHealth interventions with a fixed frequency. These trials automatically sent eHealth reminders regardless of whether or not patients took the medications. As patients become familiar with reminders, they will gradually become habitualized and generate response fatigue to the eHealth intervention, which may have a negative impact on the long-term effectiveness of interventions. Some of the trials in this review focus on real-time adherence monitoring, which only provides intervention when the patients fail to take the medicine on time, thus avoiding habitualization of reminders [60,63]. Although the 2 trials did not find a significant pooled effect of real-time reminders, it should be noted that the number of available studies limited statistical power. Future adherence intervention studies should strengthen study design in both sequence generation and blinding and should focus on real-time adherence monitoring to enhance the long-term effectiveness of eHealth interventions.

Another interesting area of this review is the effects of the eHealth interventions characteristics. As the number of available studies in some subgroups limited statistical power, the results should be considered uncertain, so we recommend that the comparison between these subgroups should be interpreted with caution. SMS did not have a significant effect on improving ART adherence in this review; however, the result was inconsistent with the result of the study by Finitsis et al [7]. This may be because SMSs are facing challenges from internet protocol–based messaging services in recent years (such as Apple's iMessage, WhatsApp, Facebook Messenger, WeChat, and Line). Therefore, the attention and use of high-cost SMSs are gradually decreasing. People's reactions to application-to-peer messaging in their daily lives have also weakened. Although previous studies have suggested that the outcomes can be improved by changing certain intervention characteristics (eg, increasing the frequency of the intervention [78] and performing 2-way communications [79]), no significant heterogeneity between these subgroups was observed in this review. Further research could use the "nudge theory" to guide the design of the eHealth interventions procedure for improving ART adherence of PLWH. The theory emphasizes that nudges are not mandatory, and their intervention design must be simple and inexpensive [80]. It was explored in previous studies that this theory had a positive impact on several behaviors, such as reducing tobacco use [81], changing adult dietary choices [82], and increasing physical activity [83]. For medication reminders, any intervention that directly asks participants about trial content should be excluded, as this would bias the participants.

Selection of Adherence Outcome Assessment Method

Although an array of methods are proposed to assess adherence, few meet the gold standards of reliability, ease of use, low cost, flexibility, and practicality. However, each method has its advantages and disadvantages. According to Lam and Fresco, subjective methods can generally explain nonadherence, whereas objective methods can more accurately measure patient adherence to medication [84]. Subjective methods have the advantage of low cost, simplicity, practicality, and flexibility. However, poor sensitivity and specificity remain an issue, and questionnaires are unreliable in terms of adherence outcomes. The patient's psychological state can also influence the accuracy of the outcomes. Outcomes are more accurate for objective

methods than those for subjective methods. However, different objective methods have variable characteristics. Although pill counting is simple and low cost, it fails to identify the medication-taking pattern. Electronic monitoring devices are only suitable for small-scale research as expensive technical support is required. Considering both accuracy and cost, pharmacy refill record is more beneficial for large numbers of research populations [85]. The assessment of biochemical outcomes can directly reflect overall treatment regimens and indirectly reflect the effectiveness of the interventions. However, these methods are expensive and intrusive. Considering the advantages and disadvantages of various outcome assessment methods, we recommend that these methods should be applied in combination in future research according to the characteristics of each study to achieve measurement purpose.

Strengths and Limitations

This review has several strengths. First, this review only includes RCTs, which are considered to enhance the methodological quality of the meta-analysis and strengthen the conclusions. Another strength is that the results of our meta-analysis indicated no influence of publication bias. In addition, subgroup analyses were performed to explore the source of between-study heterogeneity. We examined numerous moderators that significantly contribute to the design and implementation of eHealth interventions. Moreover, we performed some sensitivity analyses to detect the robustness of our results.

We found that the included studies on eHealth interventions had several limitations. First, of the 19 studies, 15 had a high or unclear risk of bias for at least one of the bias items in the methodological quality assessment. The low quality of the studies may bias the meta-analysis and reduce further pooled analysis [26]. Moreover, some of the primary outcome measures were expressed by the proportion of patients with good adherence. The level of adherence that was defined as “good” differed across the trails (thresholds were 90% in 2 trials, 95% in 4 trials, and 100% in 1 trial). Low thresholds may overestimate the effectiveness of eHealth on ART adherence [8].

Several limitations of this review should also be considered when we interpret the findings. The findings are inevitably limited by the number of studies in some moderators in the subgroup analyses that make it difficult to generalize their results. Several moderators examined in subgroup analyses may also impact each other, so they should be interpreted with caution. In addition, although we calculated Cohen *d* to standardize these measures, methodological disadvantages were observed when the adherence measures were pooled [49]. Furthermore, although the design of the RCT can provide strong evidence, it is precisely because of the rigorous randomization, blinding, quality control, and other design in the RCT that the effect of the research often deviates from the actual effect in the “real world.” Finally, we restricted the study of English language publications, and further studies across a range of ethnicities would further strengthen the findings.

Conclusions

We found that some of the eHealth interventions may be the effective method to increase the ART adherence of PLWH. The advantages of low cost, ease of access, and confidentiality make it a useful intervention tool in the PLWH. Although our analyses suggest some heterogeneity across trials, this finding is likely because of variation in the characteristics of the studies and in the definitions of outcomes among the studies. Considering that most of the trials are with small sample sizes or short-term duration, these results should be interpreted with caution. Therefore, the effectiveness of eHealth interventions in the “real world” remains uncertain.

To better identify the role of eHealth interventions in improving ART adherence of PLWH, future research needs to determine the features of eHealth interventions to better improve ART adherence along with long-term effectiveness of interventions, effectiveness of real-time adherence monitoring, enhancement of study design, and influences on biochemical outcomes. In addition, further research can try to design and implement the optimal strategy of eHealth intervention based on *nudge theory* combined with educational and psychosocial interventions.

Acknowledgments

This work is supported by the National Natural Science Foundation of China, grant numbers 71473269 and 71673301. This project is also sponsored by Liaoning BaiQianWan Talents program and Liaoning Distinguished Professor Program. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would like to thank all the authors who shared their data for use in this review. In addition, the authors are particularly grateful to the reviewers and editors whose comments were valuable and helpful for improving this study and providing significant guidance for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reviews literature search strategy.
[PDF File (Adobe PDF File), 114 KB - mhealth_v7i10e14404_app1.pdf]

Multimedia Appendix 2

Formal systematic literature search strategy.
[PDF File (Adobe PDF File), 113 KB - mhealth_v7i10e14404_app2.pdf]

Multimedia Appendix 3
Population, interventions, comparisons, outcomes and study design (PICOS) criteria for study inclusion.
[PDF File (Adobe PDF File), 115 KB - mhealth_v7i10e14404_app3.pdf]

Multimedia Appendix 4
Cochrane risk of bias quality assessment for included studies.
[PDF File (Adobe PDF File), 140 KB - mhealth_v7i10e14404_app4.pdf]

Multimedia Appendix 5
List of included studies after full-text review.
[PDF File (Adobe PDF File), 137 KB - mhealth_v7i10e14404_app5.pdf]

Multimedia Appendix 6
List of excluded studies after full-text review.
[PDF File (Adobe PDF File), 223 KB - mhealth_v7i10e14404_app6.pdf]

Multimedia Appendix 7
Study and participants characteristics of trials for principal systematic literature review.
[PDF File (Adobe PDF File), 222 KB - mhealth_v7i10e14404_app7.pdf]

Multimedia Appendix 8
Characteristics of electronic health intervention and outcome measures of trials for principal systematic literature review.
[PDF File (Adobe PDF File), 227 KB - mhealth_v7i10e14404_app8.pdf]

References

1. World Health Organization. 2019. HIV/AIDS: Key Facts URL: https://www.who.int/en/news-room/fact-sheets/detail/hiv-aids [accessed 2019-01-18] [WebCite Cache ID 77V0iXBD7]
2. The Joint United Nations Programme on HIV/AIDS (UNAIDS). 2017. 90–90–90 - An Ambitious Treatment Target to Help End the Aids Epidemic URL: http://www.unaids.org/en/resources/documents/2017/90-90-90 [accessed 2019-01-19] [WebCite Cache ID 77V18hHzF]
3. World Health Organization. 2018. HIV/AIDS: Data and Statistics URL: https://www.who.int/hiv/data/en/ [accessed 2019-01-19] [WebCite Cache ID 77V18Ulwr]
4. The Joint United Nations Programme on HIV/AIDS (UNAIDS). 2017. Ending AIDS: Progress Towards the 90–90–90 Targets URL: http://www.unaids.org/sites/default/files/media_asset/Global_AIDS_update_2017_en.pdf [accessed 2019-01-18] [WebCite Cache ID 77V14Wqtk]
5. 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 Mar 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
6. Sherr L, Lampe FC, Clucas C, Johnson M, Fisher M, Date HL, et al. Self-reported non-adherence to ART and virological outcome in a multiclinic UK study. AIDS Care 2010 Aug;22(8):939-945. [doi: 10.1080/09540121.2010.482126] [Medline: 20574863]
7. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
8. Thakkar J, Kurup R, Laba TL, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
9. Ho PM, Magid DJ, Shetterly SM, Olson KL, Maddox TM, Peterson PN, et al. Medication nonadherence is associated with a broad range of adverse outcomes in patients with coronary artery disease. Am Heart J 2008 Apr;155(4):772-779. [doi: 10.1016/j.ahj.2007.12.011] [Medline: 18371492]
10. Ferreira JL, Rodrigues R, Lança AM, de Almeida VC, Rocha SQ, Ragazzo TG, et al. Transmitted drug resistance among people living with HIV/Aids at major cities of Sao Paulo state, Brazil. Adv Virol 2013;2013:878237 [FREE Full text] [doi: 10.1155/2013/878237] [Medline: 23401688]

11. Kanters S, Park JJ, Chan K, Socias ME, Ford N, Forrest JI, et al. Interventions to improve adherence to antiretroviral therapy: a systematic review and network meta-analysis. Lancet HIV 2017 Jan;4(1):e31-e40. [doi: 10.1016/S2352-3018(16)30206-5] [Medline: 27863996]
12. 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. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
13. World Health Organization. Global Diffusion of eHealth: Making Universal Health Coverage Achievable: Report of the Third Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2016.
14. Daher J, Vijh R, Linthwaite B, Dave S, Kim J, Dheda K, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). BMJ Open 2017 Nov 3;7(11):e017604 [FREE Full text] [doi: 10.1136/bmjopen-2017-017604] [Medline: 29101138]
15. Katz IT, Ryu AE, Onuegbu AG, Psaros C, Weiser SD, Bangsberg DR, et al. Impact of HIV-related stigma on treatment adherence: systematic review and meta-synthesis. J Int Aids Soc 2013 Nov 13;16(3 Suppl 2):18640 [FREE Full text] [doi: 10.7448/IAS.16.3.18640] [Medline: 24242258]
16. Burch LS, Smith CJ, Phillips AN, Johnson MA, Lampe FC. Socioeconomic status and response to antiretroviral therapy in high-income countries: a literature review. AIDS 2016 May 15;30(8):1147-1162. [doi: 10.1097/QAD.0000000000001068] [Medline: 26919732]
17. Lehman JS, Carr MH, Nichol AJ, Ruisanchez A, Knight DW, Langford AE, et al. Prevalence and public health implications of state laws that criminalize potential HIV exposure in the United States. AIDS Behav 2014 Jun;18(6):997-1006 [FREE Full text] [doi: 10.1007/s10461-014-0724-0] [Medline: 24633716]
18. Kagee A, Remien RH, Berkman A, Hoffman S, Campos L, Swartz L. Structural barriers to ART adherence in Southern Africa: challenges and potential ways forward. Glob Public Health 2011;6(1):83-97 [FREE Full text] [doi: 10.1080/17441691003796387] [Medline: 20509066]
19. Purnomo J, Coote K, Mao L, Fan L, Gold J, Ahmad R, et al. Using ehealth to engage and retain priority populations in the HIV treatment and care cascade in the Asia-Pacific region: a systematic review of literature. BMC Infect Dis 2018 Feb 17;18(1):82 [FREE Full text] [doi: 10.1186/s12879-018-2972-5] [Medline: 29454322]
20. Tran BX, Houston S. Mobile phone-based antiretroviral adherence support in Vietnam: feasibility, patient's preference, and willingness-to-pay. AIDS Behav 2012 Oct;16(7):1988-1992. [doi: 10.1007/s10461-012-0271-5] [Medline: 22814571]
21. Rodrigues R, Bogg L, Shet A, Kumar DS, de Costa A. Mobile phones to support adherence to antiretroviral therapy: what would it cost the Indian National AIDS Control Programme? J Int AIDS Soc 2014;17:19036 [FREE Full text] [doi: 10.7448/IAS.17.1.19036] [Medline: 25186918]
22. Ericsson - A World of Communication. 2018. Mobile Subscriptions Worldwide Q3 2018 URL: https://www.ericsson.com/en/mobility-report/reports/november-2018/mobile-subscriptions-worldwide-q3-2018 [WebCite Cache ID 77V2s7M6H]
23. Statista. 2018. Number of Monthly Active WeChat Users From 2nd Quarter 2012 to 2nd Quarter 2019 (in Millions) URL: http://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/ [accessed 2019-01-21] [WebCite Cache ID 77V2weNez]
24. Alzougool B. The impact of motives for Facebook use on Facebook addiction among ordinary users in Jordan. Int J Soc Psychiatry 2018 Sep;64(6):528-535. [doi: 10.1177/0020764018784616] [Medline: 29939103]
25. Statista. 2018. Number of YouTube Users Worldwide From 2016 to 2021 (in Billions) URL: https://www.statista.com/statistics/805656/number-youtube-viewers-worldwide/ [accessed 2019-01-22] [WebCite Cache ID 77V31hrQf]
26. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
27. Rao D, Frey S, Ramaiya M. eHealth for stigma reduction efforts designed to improve engagement in care for people living with HIV. Curr HIV/AIDS Rep 2018 Dec;15(6):397-402. [doi: 10.1007/s11904-018-0414-z] [Medline: 30232579]
28. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 4;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
29. Alexander GL, McClure JB, Calvi JH, Divine GW, Stopponi MA, Rolnick SJ, MENU Choices Team. A randomized clinical trial evaluating online interventions to improve fruit and vegetable consumption. Am J Public Health 2010 Mar;100(2):319-326 [FREE Full text] [doi: 10.2105/AJPH.2008.154468] [Medline: 20019315]
30. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
31. Lima IC, Galvão MT, Alexandre HD, Lima FE, Araújo TL. Information and communication technologies for adherence to antiretroviral treatment in adults with HIV/AIDS. Int J Med Inform 2016 Aug;92:54-61. [doi: 10.1016/j.ijmedinf.2016.04.013] [Medline: 27318071]
32. Jongbloed K, Parmar S, van der Kop M, Spittal PM, Lester RT. Recent evidence for emerging digital technologies to support global HIV engagement in care. Curr HIV/AIDS Rep 2015 Dec;12(4):451-461 [FREE Full text] [doi: 10.1007/s11904-015-0291-7] [Medline: 26454756]

33. Campbell JI, Haberer JE. Cell phone-based and adherence device technologies for HIV care and treatment in resource-limited settings: recent advances. Curr HIV/AIDS Rep 2015 Dec;12(4):523-531. [doi: 10.1007/s11904-015-0282-8] [Medline: 26439917]
34. Simoni JM, Kutner BA, Horvath KJ. Opportunities and challenges of digital technology for HIV treatment and prevention. Curr HIV/AIDS Rep 2015 Dec;12(4):437-440 [FREE Full text] [doi: 10.1007/s11904-015-0289-1] [Medline: 26412082]
35. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database of Systematic Reviews 2012 Mar 14(3):CD009756 [FREE Full text] [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
36. Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews and framework for evidence transfer. BMC Health Serv Res 2015 Jan 22;15:33 [FREE Full text] [doi: 10.1186/s12913-014-0654-6] [Medline: 25609559]
37. Amankwaa I, Boateng D, Quansah DY, Akuoko CP, Evans C. Effectiveness of short message services and voice call interventions for antiretroviral therapy adherence and other outcomes: a systematic review and meta-analysis. PLoS One 2018;13(9):e0204091 [FREE Full text] [doi: 10.1371/journal.pone.0204091] [Medline: 30240417]
38. Muessig KE, LeGrand S, Horvath KJ, Bauermeister JA, Hightow-Weidman LB. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441 [FREE Full text] [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
39. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015 Jun 2;162(11):777-784. [doi: 10.7326/M14-2385] [Medline: 26030634]
40. Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: Wiley; 2008.
41. Spaan P, van Luenen S, Garnefski N, Kraaij V. Psychosocial interventions enhance HIV medication adherence: A systematic review and meta-analysis. J Health Psychol 2018 Feb 01:1359105318755545. [Epub ahead of print]. [doi: 10.1177/1359105318755545] [Medline: 29417851]
42. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res 2018 Jun;27(6):1785-1805. [doi: 10.1177/0962280216669183] [Medline: 27683581]
43. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
44. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
45. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
46. Cooper H, Hedges LV. The Handbook Of Research Synthesis. New York: Russell Sage Foundation; 1997.
47. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Br Med J 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
48. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341X.2000.00455.x] [Medline: 10877304]
49. Tao D, Xie L, Wang T, Wang T. A meta-analysis of the use of electronic reminders for patient adherence to medication in chronic disease care. J Telemed Telecare 2015 Jan;21(1):3-13. [doi: 10.1177/1357633X14541041] [Medline: 25147178]
50. Safren SA, Hendriksen ES, Desousa N, Boswell SL, Mayer KH. Use of an on-line pager system to increase adherence to antiretroviral medications. AIDS Care 2003 Dec;15(6):787-793. [doi: 10.1080/09540120310001618630] [Medline: 14617500]
51. Reynolds NR, Testa MA, Su M, Chesney MA, Neidig JL, Frank I, AIDS Clinical Trials Group 731384 Teams. Telephone support to improve antiretroviral medication adherence: a multisite, randomized controlled trial. J Acquir Immune Defic Syndr 2008 Jan 1;47(1):62-68. [doi: 10.1097/QAI.0b013e3181582d54] [Medline: 17891043]
52. Simoni JM, Huh D, Frick PA, Pearson CR, Andrasik MP, Dunbar PJ, et al. Peer support and pager messaging to promote antiretroviral modifying therapy in Seattle: a randomized controlled trial. J Acquir Immune Defic Syndr 2009 Dec 1;52(4):465-473 [FREE Full text] [doi: 10.1097/QAI.0b013e3181b9300c] [Medline: 19911481]
53. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
54. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
55. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills EJ, Smieja M, et al. The Cameroon Mobile Phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. PLoS One 2012;7(12):e46909 [FREE Full text] [doi: 10.1371/journal.pone.0046909] [Medline: 23236345]

56. da Costa TM, Barbosa BJ, Costa DA, Sigulem D, de Fátima MH, Filho AC, et al. Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages. Int J Med Inform 2012 Apr;81(4):257-269 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.10.002] [Medline: 22296762]
57. Hersch RK, Cook RF, Billings DW, Kaplan S, Murray D, Safren S, et al. Test of a web-based program to improve adherence to HIV medications. AIDS Behav 2013 Nov;17(9):2963-2976 [FREE Full text] [doi: 10.1007/s10461-013-0535-8] [Medline: 23760634]
58. Shet A, de Costa A, Kumarasamy N, Rodrigues R, Rewari BB, Ashorn P, HIVIND Study Team. Effect of mobile telephone reminders on treatment outcome in HIV: evidence from a randomised controlled trial in India. Br Med J 2014 Oct 24;349:g5978 [FREE Full text] [doi: 10.1136/bmj.g5978] [Medline: 25742320]
59. Robbins GK, Testa MA, Su M, Safren SA, Morse G, Lammert S, et al. Site nurse-initiated adherence and symptom support telephone calls for HIV-positive individuals starting antiretroviral therapy, ACTG 5031: substudy of ACTG 384. HIV Clin Trials 2013;14(5):235-253 [FREE Full text] [doi: 10.1310/hct1405-235] [Medline: 24144900]
60. Sabin LL, DeSilva MB, Gill CJ, Zhong L, Vian T, Xie W, et al. Improving adherence to antiretroviral therapy with triggered real-time text message reminders: the China adherence through technology study. J Acquir Immune Defic Syndr 2015 Aug 15;69(5):551-559 [FREE Full text] [doi: 10.1097/QAI.0000000000000651] [Medline: 25886927]
61. Ingersoll KS, Dillingham RA, Hettema JE, Conaway M, Freeman J, Reynolds G, et al. Pilot RCT of bidirectional text messaging for ART adherence among nonurban substance users with HIV. Health Psychol 2015 Dec;34S:1305-1315 [FREE Full text] [doi: 10.1037/hea0000295] [Medline: 26651472]
62. Belzer ME, Naar-King S, Olson J, Sarr M, Thornton S, Kahana SY, Adolescent Medicine Trials Network for HIV/AIDS Interventions. The use of cell phone support for non-adherent HIV-infected youth and young adults: an initial randomized and controlled intervention trial. AIDS Behav 2014 Apr;18(4):686-696 [FREE Full text] [doi: 10.1007/s10461-013-0661-3] [Medline: 24271347]
63. Orrell C, Cohen K, Mauff K, Bangsberg DR, Maartens G, Wood R. A randomized controlled trial of real-time electronic adherence monitoring with text message dosing reminders in people starting first-line antiretroviral therapy. J Acquir Immune Defic Syndr 2015 Dec 15;70(5):495-502. [doi: 10.1097/QAI.0000000000000770] [Medline: 26218411]
64. Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A randomized controlled trial of personalized text message reminders to promote medication adherence among HIV-positive adolescents and young adults. AIDS Behav 2016 May;20(5):1049-1059 [FREE Full text] [doi: 10.1007/s10461-015-1192-x] [Medline: 26362167]
65. Ruan Y, Xiao X, Chen J, Li X, Williams AB, Wang H. Acceptability and efficacy of interactive short message service intervention in improving HIV medication adherence in Chinese antiretroviral treatment-naïve individuals. Patient Prefer Adherence 2017;11:221-228 [FREE Full text] [doi: 10.2147/PPA.S120003] [Medline: 28228652]
66. Reid MJ, Steenhoff AP, Thompson J, Gabaitiri L, Cary MS, Steele K, et al. Evaluation of the effect of cellular SMS reminders on consistency of antiretroviral therapy pharmacy pickups in HIV-infected adults in Botswana: a randomized controlled trial. Health Psychol Behav Med 2017;5(1):101-109 [FREE Full text] [doi: 10.1080/21642850.2016.1271333] [Medline: 28966882]
67. Abdulrahman SA, Rampal L, Ibrahim F, Radhakrishnan AP, Shahar HK, Othman N. Mobile phone reminders and peer counseling improve adherence and treatment outcomes of patients on ART in Malaysia: a randomized clinical trial. PLoS One 2017;12(5):e0177698 [FREE Full text] [doi: 10.1371/journal.pone.0177698] [Medline: 28520768]
68. Linnemayr S, Huang H, Luoto J, Kambugu A, Thirumurthy H, Haberer JE, et al. Text messaging for improving antiretroviral therapy adherence: no effects after 1 year in a randomized controlled trial among adolescents and young adults. Am J Public Health 2017 Dec;107(12):1944-1950. [doi: 10.2105/AJPH.2017.304089] [Medline: 29048966]
69. Conn VS, Ruppar TM. Medication adherence outcomes of 771 intervention trials: systematic review and meta-analysis. Prev Med 2017 Jun;99:269-276 [FREE Full text] [doi: 10.1016/j.ypmed.2017.03.008] [Medline: 28315760]
70. Lowther K, Selman L, Harding R, Higginson IJ. Experience of persistent psychological symptoms and perceived stigma among people with HIV on antiretroviral therapy (ART): a systematic review. Int J Nurs Stud 2014 Aug;51(8):1171-1189. [doi: 10.1016/j.ijnurstu.2014.01.015] [Medline: 24602830]
71. Shubber Z, Mills EJ, Nachega JB, Vreeman R, Freitas M, Bock P, et al. Patient-reported barriers to adherence to antiretroviral therapy: a systematic review and meta-analysis. PLoS Med 2016 Nov;13(11):e1002183 [FREE Full text] [doi: 10.1371/journal.pmed.1002183] [Medline: 27898679]
72. Mohammed MA, Moles RJ, Chen TF. Medication-related burden and patients' lived experience with medicine: a systematic review and metasynthesis of qualitative studies. BMJ Open 2016 Mar 2;6(2):e010035 [FREE Full text] [doi: 10.1136/bmjopen-2015-010035] [Medline: 26839015]
73. Sidebottom D, Ekström AM, Strömdahl S. A systematic review of adherence to oral pre-exposure prophylaxis for HIV - how can we improve uptake and adherence? BMC Infect Dis 2018 Nov 16;18(1):581 [FREE Full text] [doi: 10.1186/s12879-018-3463-4] [Medline: 30445925]
74. Burch LS, Smith CJ, Anderson J, Sherr L, Rodger AJ, O'Connell R, et al. Socioeconomic status and treatment outcomes for individuals with HIV on antiretroviral treatment in the UK: cross-sectional and longitudinal analyses. Lancet Public Health 2016 Nov;1(1):e26-e36 [FREE Full text] [doi: 10.1016/S2468-2667(16)30002-0] [Medline: 28299369]

75. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 2001 Oct;54(10):1046-1055. [doi: 10.1016/S0895-4356(01)00377-8] [Medline: 11576817]
76. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001 Dec 4;135(11):982-989. [doi: 10.7326/0003-4819-135-11-200112040-00010] [Medline: 11730399]
77. Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. Critical Care 2013 Jan 09;17(1):1-9 [FREE Full text] [doi: 10.1186/cc11919] [Medline: 23302257]
78. Vervloet M, Linn AJ, van Weert JC, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
79. Wald DS, Butt S, Bestwick JP. One-way versus two-way text messaging on improving medication adherence: meta-analysis of randomized trials. Am J Med 2015 Oct;128(10):1139.e1-1139.e5. [doi: 10.1016/j.amjmed.2015.05.035] [Medline: 26087045]
80. Thaler RH, Sunstein CR. Nudge: Improving Decisions About Health, Wealth, and Happiness. New Haven, CT: Yale University Press; 2008.
81. Hall MG, Marteau TM, Sunstein CR, Ribisl KM, Noar SM, Orlan EN, et al. Public support for pictorial warnings on cigarette packs: an experimental study of US smokers. J Behav Med 2018 Jun;41(3):398-405 [FREE Full text] [doi: 10.1007/s10865-018-9910-2] [Medline: 29411272]
82. Arno A, Thomas S. The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. BMC Public Health 2016 Jul 30;16:676 [FREE Full text] [doi: 10.1186/s12889-016-3272-x] [Medline: 27475752]
83. Harrison JD, Jones JM, Small DS, Rareshide CA, Szwartz G, Steier D, et al. Social incentives to encourage physical activity and understand predictors (STEP UP): design and rationale of a randomized trial among overweight and obese adults across the United States. Contemp Clin Trials 2019 May;80:55-60. [doi: 10.1016/j.cct.2019.04.001] [Medline: 30954675]
84. Lam WY, Fresco P. Medication adherence measures: an overview. Biomed Res Int 2015;2015:217047 [FREE Full text] [doi: 10.1155/2015/217047] [Medline: 26539470]
85. Sangeda RZ, Mosha F, Prosperi M, Aboud S, Vercauteren J, Camacho RJ, et al. Pharmacy refill adherence outperforms self-reported methods in predicting HIV therapy outcome in resource-limited settings. BMC Public Health 2014 Oct 4;14:1035 [FREE Full text] [doi: 10.1186/1471-2458-14-1035] [Medline: 25280535]

Abbreviations

ART: antiretroviral therapy
CA: control arm
CD4$^+$: cluster of differentiation 4$^+$
EAMD: electronic adherence monitoring device
eHealth: electronic health
IA: intervention arm
LMICs: low- and middle-income countries
mHealth: mobile health
OR: odds ratio
PICOS: populations, interventions, comparisons, outcomes, and study design
PLWH: people living with HIV
RCTs: randomized controlled trials
SMS: short message service
VF: virological failure
VS: viral suppression
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.04.19; peer-reviewed by J Mitchell, A Thomas, C Conn; comments to author 01.06.19; revised version received 25.07.19; accepted 18.08.19; published 16.10.19.

Please cite as:
Wang Z, Zhu Y, Cui L, Qu B
Electronic Health Interventions to Improve Adherence to Antiretroviral Therapy in People Living With HIV: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2019;7(10):e14404
URL: https://mhealth.jmir.org/2019/10/e14404
doi:10.2196/14404
PMID:31621641

Original Paper

The Effect of an App for Day-to-Day Postoperative Care Education on Patients With Total Knee Replacement: Randomized Controlled Trial

Thomas Timmers[1,2], MSc; Loes Janssen[3], PhD; Walter van der Weegen[4], PhD; Dirk Das[4], MD; Willem-Jan Marijnissen[5], MD, PhD; Gerjon Hannink[6], PhD; Babette C van der Zwaard[7], PhD; Adriaan Plat[7], MD; Bregje Thomassen[8], PhD; Jan-Willem Swen[8], MD; Rudolf B Kool[2], MD, PhD; Frederik Okke Lambers Heerspink[9], MD, PhD

[1]Interactive Studios, Rosmalen, Netherlands

[2]Radboud university medical center, Radboud Institute for Health Sciences, IQ healthcare, Nijmegen, Netherlands

[3]Máxima MC, Veldhoven, Netherlands

[4]Sint Anna Hospital, Geldrop, Netherlands

[5]Albert Schweitzer Hospital, Dordrecht, Netherlands

[6]Radboud university medical center, Nijmegen, Netherlands

[7]Jeroen Bosch Hospital, 's-Hertogenbosch, Netherlands

[8]Haaglanden Medical Center, 's-Gravenhage, Netherlands

[9]VieCuri Medical Centre, Venlo, Netherlands

Corresponding Author:
Thomas Timmers, MSc
Interactive Studios
Huisbergenweg 6
Rosmalen
Netherlands
Phone: 31 736446069
Email: thomas@interactivestudios.nl

Abstract

Background: Patients who undergo primary Total Knee Replacement surgery (TKR) are often discharged within 1-3 days after surgery. With this relatively short length of hospital stay, a patient's self-management is a crucial factor in optimizing the outcome of their treatment. In the case of TKR, self-management primarily involves adequate pain management, followed by physiotherapy exercises and daily self-care activities. Patients are educated on all these topics by hospital staff upon discharge from the hospital but often struggle to comprehend this information due to its quantity, complexity, and the passive mode of communication used to convey it.

Objective: This study primarily aims to determine whether actively educating TKR patients with timely, day-to-day postoperative care information through an app could lead to a decrease in their level of pain compared to those who only receive standard information about their recovery through the app. In addition, physical functioning, quality of life, ability to perform physiotherapy exercises and daily self-care activities, satisfaction with information, perceived involvement by the hospital, and health care consumption were also assessed.

Methods: A multicenter randomized controlled trial was performed in five Dutch hospitals. In total, 213 patients who had undergone elective, primary, unilateral TKR participated. All patients had access to an app for their smartphone and tablet to guide them after discharge. The intervention group could unlock day-to-day information by entering a personal code. The control group only received weekly, basic information. Primary (level of pain) and secondary outcomes (physical functioning, quality of life, ability to perform physiotherapy exercises and activities of daily self-care, satisfaction with information, perceived involvement by the hospital, and health care consumption) were measured using self-reported online questionnaires. All outcomes were measured weekly in the four weeks after discharge, except for physical functioning and quality of life, which were measured at baseline and at four weeks after discharge. Data was analyzed using Student *t* tests, chi-square tests, and linear mixed models for repeated measures.

Results: In total, 114 patients were enrolled in the intervention group (IG) and 99 in the control group (CG). Four weeks after discharge, patients in the IG performed significantly better than patients in the CG on all dimensions of pain: pain at rest (mean 3.45 vs mean 4.59; *P*=.001), pain during activity (mean 3.99 vs mean 5.08; *P*<.001) and pain at night (mean 4.18 vs mean 5.21; *P*=.003). Additionally, significant differences were demonstrated in favor of the intervention group for all secondary outcomes.

Conclusions: In the four weeks following TKR, the active and day-to-day education of patients via the app significantly decreased their level of pain and improved their physical functioning, quality of life, ability to perform physiotherapy exercises and activities of daily self-care, satisfaction with information, perceived involvement by the hospital, and health care consumption compared to standard patient education. Given the rising number of TKR patients and the increased emphasis on self-management, we suggest using an app with timely postoperative care education as a standard part of care.

Trial Registration: Netherlands Trial Register NTR7182; https://www.trialregister.nl/trial/6992

(JMIR Mhealth Uhealth 2019;7(10):e15323) doi:10.2196/15323

KEYWORDS

patient education; postoperative care; smartphone; self-management; ehealth; telemedicine

Introduction

Background

Osteoarthritis of the knee is one of the leading causes of disability among adults aged 65 years old and over [1]. Complaints often result in a loss of productivity and reduced quality of life [2,3]. Total Knee Replacement (TKR) is considered the best available treatment option when conservative options have failed, resulting in a postoperative relief of pain, increased functional outcomes, and a patient satisfaction of more than 80% [4]. TKR is one of the most commonly performed orthopedic procedures internationally and in the United States alone its rate of occurrence is estimated to increase to more than 3.4 million procedures by the year 2030 [3].

In the last decade, TKR fast-track (or enhanced recovery) pathways have been implemented in many hospitals. These pathways imply a combination of patient education, multimodal analgesia, early mobilization, fluid management and nutrition optimization [5,6]. These fast-track pathways have led to a substantial decrease in patients' length of hospital stay following TKR, which reduced, on average, from 10-11 days in 2000 to 2-4 days in 2013 [7]. Nowadays, TKR is sometimes even performed as a one-day procedure in which patients do not need to stay overnight in the hospital [8]. Time to full recovery after TKR takes an average of 6-12 months.

Because of the shortened length of stay, patients' self-management has become a crucial factor in optimizing their health outcomes. According to the World Health Organization, postoperative self-management is the ability of individuals, families, and communities to cope with illness, with or without the support of a health care provider [9]. In the case of TKR, postoperative self-management primarily involves controlling the level of pain, followed by performing physiotherapy exercises and daily self-care activities [10-13]. Next to the available information in brochures, hospital staff educate patients on these topics to prepare them on how best to manage their new situation when back at home.

Patients often struggle to comprehend this information due to its quantity, complexity, and the passive mode of communication used to convey it [14,15]. This leads to a limited amount of knowledge and confidence regarding self-management [16,17], which is a predictor for lower adherence rates [18]. Additionally, patients' pain-related fear of movement, also referred to as kinesiophobia, is known to negatively affect TKR early outcomes [19]. This leads to lower rates of satisfaction, as patients feel the discharge process is rushed and that they are no longer being cared for until the next follow-up, which is usually six to eight weeks after discharge [16].

Electronic health (eHealth) or mobile health (mHealth) offer a potentially powerful means for active patient education and behavioral change reinforcement [20]. A 2018 review on perioperative eHealth interventions has demonstrated a positive effect upon the postoperative course in patients who are undergoing orthopedic and cardiac surgery [21]. The interventions described in the review range from an educational website to the facilitation of telemonitoring and teleconsultation. However, evidence concerning the use of smartphone or tablet apps is scarce. This is surprising given the number of people that own a mobile device, coupled with the increase in availability and usage of medical mobile applications [22,23]. Additionally, smartphones and tablets possess the unique ability to receive push-notifications that can be used to actively inform patients at times when the information becomes relevant for them. These notifications can provide patients with personalized guidance at various stages of their patient journey.

Objectives

The aim of this randomized controlled trial was to investigate the effect of an interactive app on patients' level of pain, physical functioning, quality of life, satisfaction, and health care consumption in the first four weeks of recovery after TKR. We hypothesized that, compared to standard practices of patient education, providing patients with timely, day-to-day information via an app would have a positive effect on all outcomes.

The primary outcome of the study was patients' level of pain during the first four weeks after discharge. The secondary outcomes of the study were physical functioning, quality of life, patients' ability to perform physiotherapy exercises and daily self-care activities, satisfaction with the information provided, perceived level of postoperative involvement by the hospital,

and health care consumption. All these outcomes were measured by means of self-reported, online questionnaires.

Methods

Study Design

A total of five Dutch hospitals (four nonacademic teaching hospitals and one general hospital) participated in the study. Between May and December 2018, patients scheduled for elective, primary, unilateral TKR were invited to participate in a surgeon-blinded, randomized controlled trial. The study focused on the four-week period following discharge from the hospital and assessed the effectiveness of an interactive app compared to the standard of care in a parallel group design with equal allocation ratio. No changes were made to the design after the study was commenced. We followed the Consolidated Standards of Reporting Trials (CONSORT) guidelines and the CONSORT EHEALTH Checklist [24,25].

Informed Consent and Ethical Considerations

Consent was gained via hospital staff who contacted patients by phone to ask them to consider participation in the study about two weeks prior to their TKR. Patients willing to participate received an email with all the necessary study information required for informed consent. Patients were offered at least two days to reflect on the information. In case of any questions, patients could contact the local research coordinator by phone or by email. Patients gave consent by signing an online informed consent form. There were no indicators of substantial risk as a function of participating in this study. The study was registered at the Netherlands Trial Registry (NTR), with the reference number 6992. The study was approved by the Institutional Review Board of the Maxima Medical Centre (Eindhoven, The Netherlands) with the reference number N17.158.

Participant Selection

Patients scheduled for elective, primary, unilateral TKR and aged 40 years and above were eligible for inclusion. Additionally, participants were required to be fluent in Dutch and in the possession of an email address and smartphone or tablet.

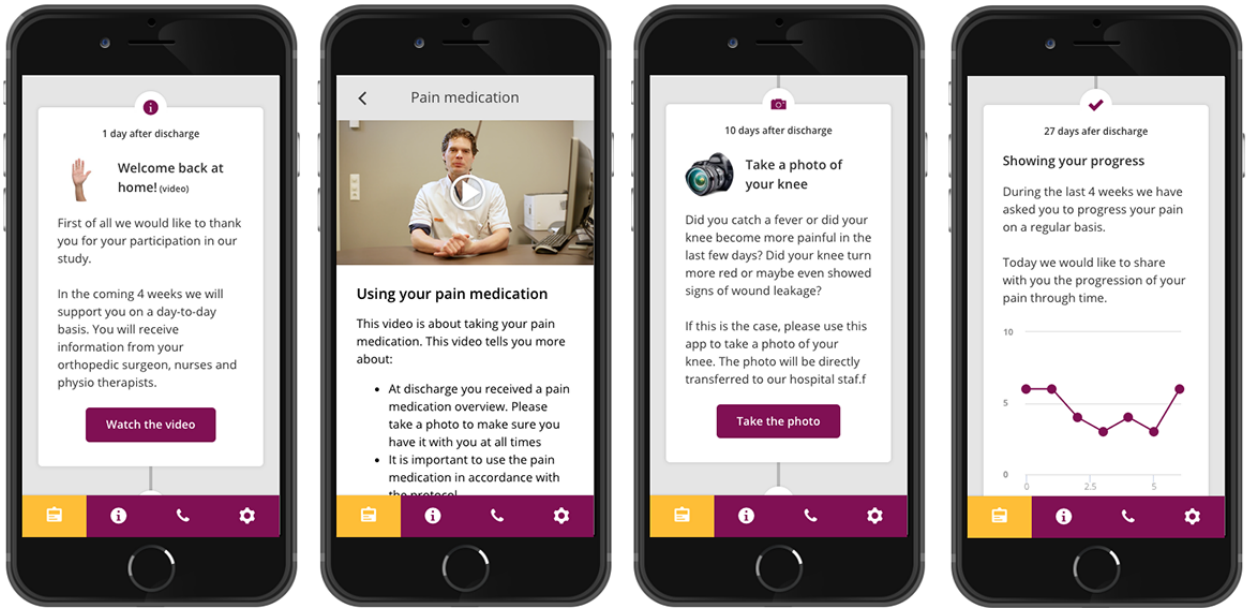
Intervention

The Patient Journey App (Interactive Studios, Rosmalen, The Netherlands) was used as the intervention to provide information to both patient groups. All patients had access to the app to guide them after discharge. The control group only received basic information about the recovery process about two times per week. The intervention group could unlock day-to-day information by entering a personal code. Participants in the intervention group received this personal code by email after completing the baseline questionnaire.

All patients in the intervention group received the same information via the app. However, aligning the timing of the information to the individual patient's phase of recovery gave it a personalized character. Patient's date of discharge was used to assure the timing of the information was correct. Push notifications were used to actively alert patients about information being available. The timing of the push notifications was configured per information item (eg, information about pain medication was available 1 day after discharge at 11 am and information about physiotherapy exercises was available 2 days after discharge at 2 pm).

The text, photos, and videos that patients in the intervention group received were developed specifically for this trial and composed and based upon interviews with orthopedic surgeons, physician assistants, nurses, and physiotherapists from participating hospitals. Furthermore, electronic health records of 50 patients who had previously undergone TKR were checked to determine for what reasons they had contacted the hospital. Based on this information, an interactive timeline was developed (Figure 1). This information was not available to patients in the control group. All information on the timeline was presented in Dutch.

Figure 1. Examples of the interactive app used as an intervention, translated from Dutch (language used in the study) to English. From left to right: the welcoming of patients to the app, video and text information about medication usage, an invitation to send a photo of the wound (in case of fever, increased levels of pain or wound leakage) and a patient-reported pain score progress tracker.



Information in the app was tailored for each hospital and based on existing protocols. During the 28-day period after discharge, every patient in the intervention group received over 30 notifications with supporting information, related to topics such as pain, physiotherapy exercises, wound care, and daily self-care activities (see Multimedia Appendix 1). Apart from the information being available in the app at the timeline, it was also available as an alphabetically ordered reference within the app. Additionally, patients were requested to enter their pain scores on a weekly basis and were able to check their results within an interactive graph every week. Patients also had the opportunity to upload a photo of the wound in case of fever and an increase in pain or wound leakage. Prior to study initiation, six TKR patients at two of the participating hospitals were interviewed to generally assess the usefulness and usability of the app. They reported that the app would be very useful and had no suggestions for changes. After the study, all the content developed for the intervention was provided to the participating hospitals, allowing them to partially offer it to their patients as the new standard of care.

Study Outcomes

As the primary outcome, the patients' ability to manage their pain was assessed in the four weeks following discharge from the hospital. As secondary outcomes, we assessed patients' physical functioning and quality of life at baseline and four weeks after discharge. Additionally, patients' ability to perform physiotherapy exercises and daily self-care activities, satisfaction with the information provided, perceived level of postoperative involvement by the hospital, and health care consumption were assessed weekly during the four weeks following discharge (Table 1). Finally, data on app usage was continuously captured to get a better understanding of how the app was being used over time, the type of information that patients consulted, and the videos they watched.

Table 1. Overview of used questionnaires per outcome.

Outcome	Questionnaire
Level of pain[a]	Three questions concerning pain while at rest, during activity and during the night. NRS[b] scores were used to measure the outcome, ranging from 0 (no pain at all) to 10 (worst pain imaginable).
Physical functioning[a]	The Knee injury and Osteoarthritis Outcome Score short form [26] involves 7 multiple choice questions indicating functional limitations, ranging from 0 (minimal limitations) to 100 (maximal limitations).
Quality of life[a]	EuroQol 3-level questionnaire, EQ-5D-3. Measuring 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety depression) on a 3-level scale (no problems, some problems, extreme problems). Additionally, a VAS[c] was used to assess patients self-rated health. VAS endpoints range from 0 (Worst imaginable state) to 100 (Best imaginable state) [27].
Physiotherapy exercises (developed for this study)	One question concerning patients' ability to perform their physiotherapy exercises. NRS score was used to measure the outcome, ranging from 0 (not capable at all) to 10 (perfectly capable).
Daily self-care activities (developed for this study)	One question concerning patients' ability to perform their daily self-care activities. NRS score was used to measure the outcome, ranging from 0 (not capable at all) to 10 (perfectly capable).
Satisfaction with information (developed for this study)	One question concerning patients' satisfaction about the information they received from the hospital after discharge. NRS score was used to measure the outcome, ranging from 0 (not satisfied at all) to 10 (very much satisfied).
Perceived involvement (developed for this study)	One question concerning the patient-perceived level of involvement by the hospital postoperatively. NRS score was used to measure the outcome, ranging from 0 (not satisfied at all) to 10 (very much satisfied).
Health care consumption (developed for this study)	Three dichotomized questions concerning health care consumption at the hospital, the general practitioner and home-care organization. When a patient indicated they had contacted one of these organizations, the outcome was assessed (ie, no action required or follow-up action required). If follow-up was required, that could entail things like a consultation or referral to another organization.

[a]These questionnaires are part of the Patient Reported Outcome Measures guidelines of the Dutch Orthopedic Association (Nederlandse Orthopaedische Vereniging, NOV) for TKR.

[b]NRS: numeric rating scale.

[c]VAS: visual analogue scale.

Some questions were developed especially for this study. These questionnaires were checked by surgeons and researchers from participating hospitals (see Multimedia Appendix 2). Additionally, a specialist organization on using acceptable language (Bureau Beter Taal, Beusichem, The Netherlands) reviewed and revised the questionnaires to assure readability for about 95% of the Dutch population (language level B1). Finally, the questionnaires were tested for usability by six patients in two of the participating hospitals prior to study initiation, which lead to no additional changes.

Study outcomes were measured five times in total: at baseline and on a weekly basis in the four weeks after discharge (Table 2). The baseline measurement was taken directly after patients were first included in the study. Follow-up questionnaires were sent to both groups at the end of each postoperative week, allowing patients to reflect on each previous week. Per measurement, a maximum of two email reminders were sent in case a patient would not respond. To minimize the risk of recall bias, patients only had a four-day time period to complete the questionnaires for each measurement. All outcome data was self-reported and collected using an online system. Patients who either missed the baseline measurement or more than two of their follow-up questionnaires were registered as lost to follow-up. These patients were not included in the analysis.

Table 2. Overview of outcomes per time point measurement.

Baseline	Follow-up week 1, 2 and 3	Follow-up week 4
Patient characteristics	_[a]	–
Physical functioning	–	Physical functioning
Quality of life	–	Quality of life
–	Level of pain	Level of pain
–	Physiotherapy exercises	Physiotherapy exercises
–	Daily self-care activities	Daily self-care activities
–	Satisfaction with information	Satisfaction with information
–	Perceived involvement	Perceived involvement
–	Health care consumption	Health care consumption

[a]Not applicable.

Sample Size

The sample size calculation was based on the 2018 Hardt et al study [28] in which an app was used postoperatively to educate patients on physiotherapy exercises and pain management. In this study a 1.0-point difference in the level of pain (on a numeric rating scale [NRS] of 0-10) was found in favor of the intervention group. Since this intervention was used in a hospital setting with nurses controlling for the right application of the intervention, we expect the effect or our intervention to be lower.

We performed an *a priori* sample size calculation (alpha=.05, [1−beta] =.90), based on a repeated measures analysis of variance (ANOVA) (groups=2, measurements=4) and an effect size of 0.25 that was based on a between-groups difference of 0.5 (SD 1.0). This calculation resulted in a minimum requirement of 78 patients in each study arm. Adding an expected loss to follow-up of 20% led to a total sample size of 190 patients. The sample size calculation was performed using G*Power, version 3.1 (Universität Düsseldorf, Düsseldorf, Germany).

Randomization

Patients were randomized by a computer to either a control or intervention group. Randomization was performed without block or stratification restrictions. After being allocated to one of the groups, patients received an email that included the link to the online informed consent form and the baseline questionnaire.

Statistical Methods

For our primary analysis we used an intention-to-treat approach including all randomized patients. Normally distributed continuous variables (eg, physical functioning and quality of life) were presented as a mean value with an SD and were statistically compared between the groups using independent Student two-tailed *t* tests. Nonnormally distributed variables were presented as a median value with the interquartile range. Categorical variables (eg, health care consumption) were presented as number and percentage and compared between groups using Chi-square tests. Linear mixed models for repeated measures were used to estimate the effect of the use of the intervention, using primary and secondary outcomes (ie, level of pain, performing physiotherapy exercises, performing activities of daily self-care, satisfaction with information, and patient-perceived involvement by the hospital) as dependent variables and intervention group, time, and the interaction between time and intervention group as fixed effect variables. Patient ID and location were used as random effect variables. Missing data were not replaced in any type of analysis.

Patients' level of education was split into two groups for the purpose of analysis: group 1 (none, elementary school, and secondary [vocational] education) and group 2 (higher secondary education, preuniversity education, and university [of applied science]). $P \leq .05$ was assumed to indicate a significant difference. *P* values between .05 and .10 were assumed to indicate a trend. As per protocol, analysis was performed to examine the robustness of our results by specifically analyzing the results of patients in the intervention group that downloaded and used the app. All data was analyzed using SPSS version 25.0, (IBM, Armonk, United States), except for the linear mixed model analysis which was executed using R, version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Study Sample

Between May and December 2018, a total of 262 eligible patients were willing to participate in the study. A total of 41 patients (15.6%) withdrew from the study without completing the baseline questionnaire (for reasons unknown). One patient (0.4%) only completed the baseline questionnaire, and an additional two patients (0.8%) did not complete more than two follow-up questionnaires. In total, 114 patients were actively enrolled in the intervention group and 99 patients in the control group. In the intervention group, 93 patients downloaded and used the app (Figure 2).

Baseline characteristics of the study population were largely similar between groups (Table 3).

Figure 2. Patient flow diagram.

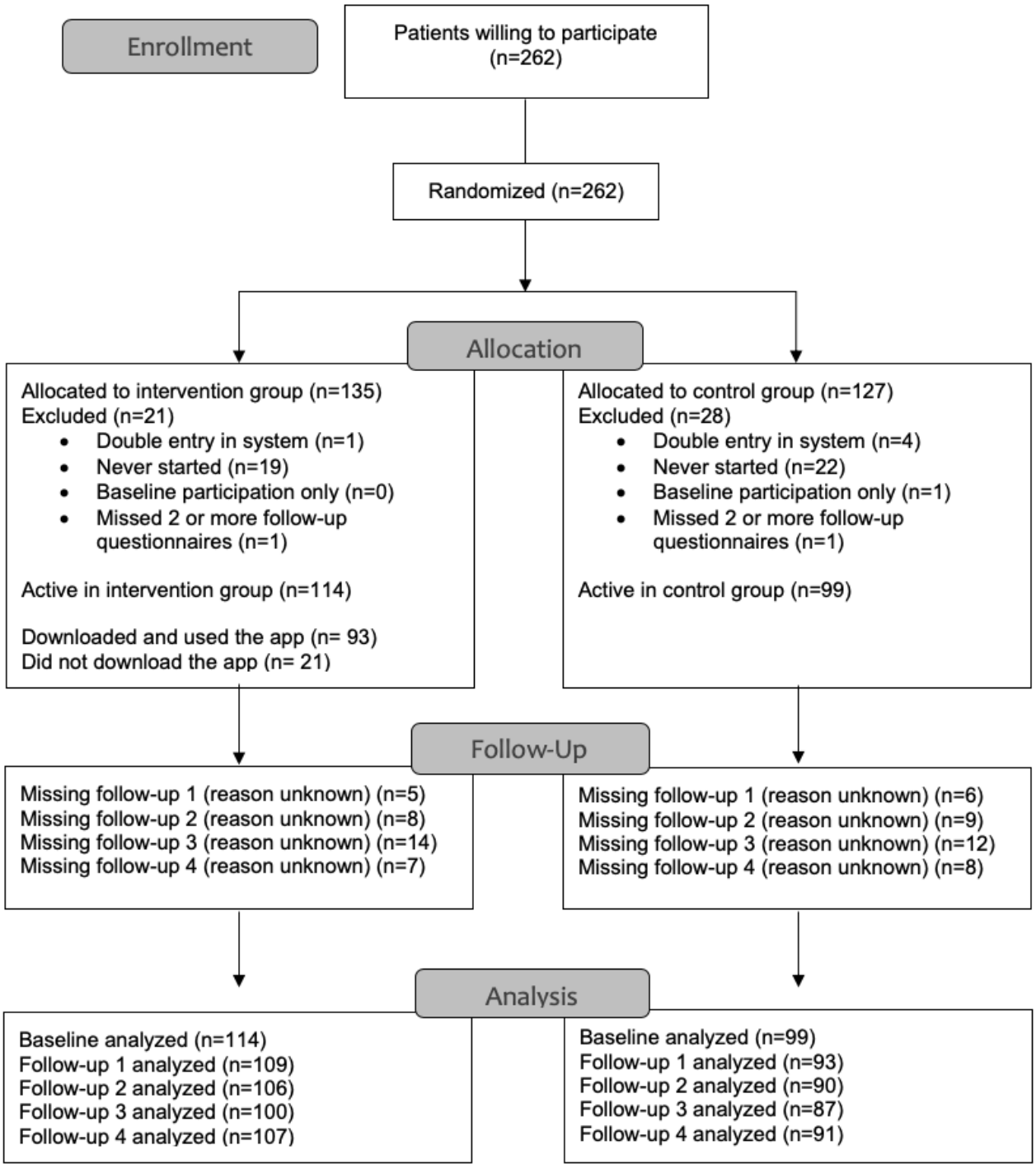


Table 3. Patient characteristics.

Characteristics	Intervention group (n=114)	Control group (n=99)
Sex, n (%)		
Male	40 (35.1)	39 (39.4)
Female	74 (64.9)	60 (60.6)
Age (years), mean (SD)	64.74 (7.57)	65.63 (7.90)
Education, n (%)		
Group 1	52 (46.0)	47 (47.5)
Group 2	61 (54.0)	52 (52.5)
Home-situation, n (%)		
Living alone	27 (23.7)	21 (21.2)
Living together	87 (76.3)	78 (78.8)
KOOS PS[a], mean (SD)	45.27 (12.71)	44.05 (11.10)
Quality of Life		
EQ-5D-3, mean (SD)	0.67 (0.21)	0.65 (0.24)
EQ-5D-3 VAS[b], mean (SD)	69.59 (16.38)	67.42 (19.56)

[a]KOOS PS: Knee injury and Osteoarthritis Outcome Score short form.

[b]VAS: visual analogue scale.

Primary Outcomes

Level of pain

In both the intervention and control group, patients' levels of pain decreased on all three dimensions of pain during the first four weeks after discharge (Figure 3, Figure 4, and Figure 5). Patients in the intervention group performed better on each of the dimensions from the second week onwards (Table 4).

Figure 3. Pain at rest (measured weekly in the first four weeks after discharge, 95% CI).

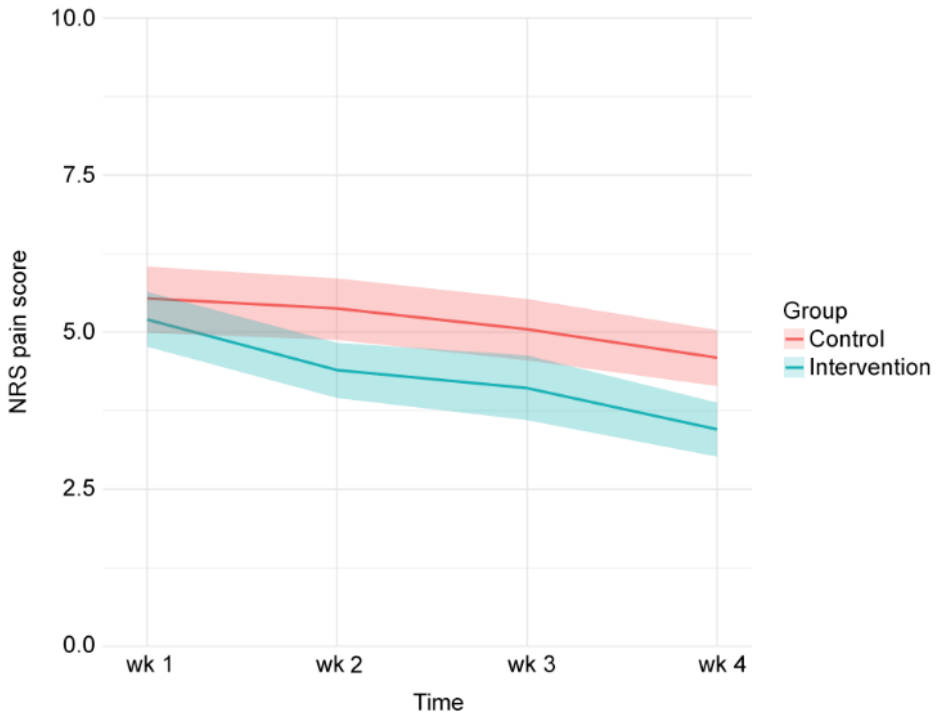


Figure 4. Pain during activity (measured weekly in the first four weeks after discharge, 95% CI).

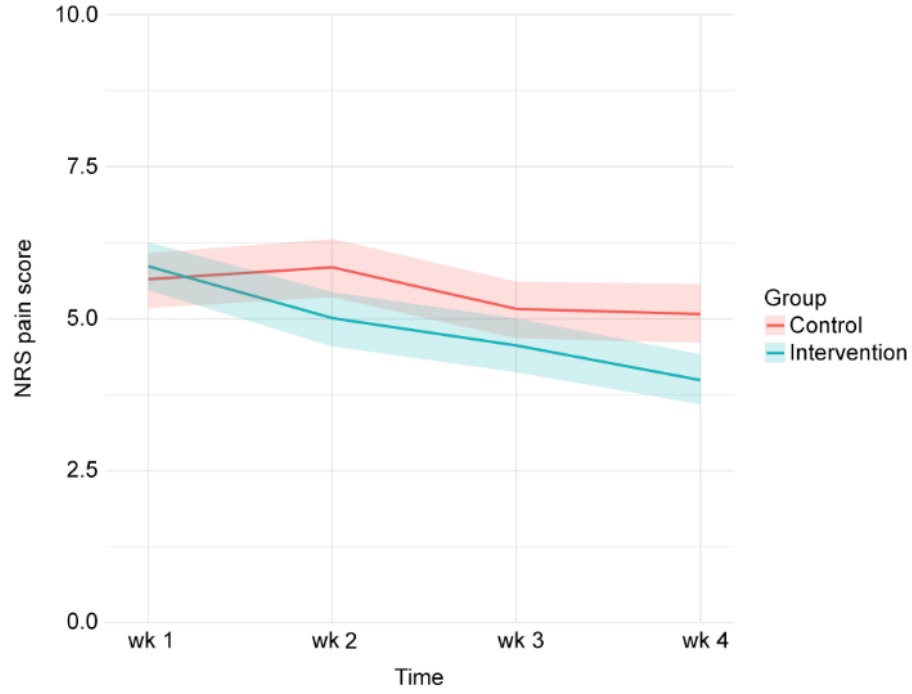


Figure 5. Pain at night (measured weekly in the first four weeks after discharge, 95% CI).

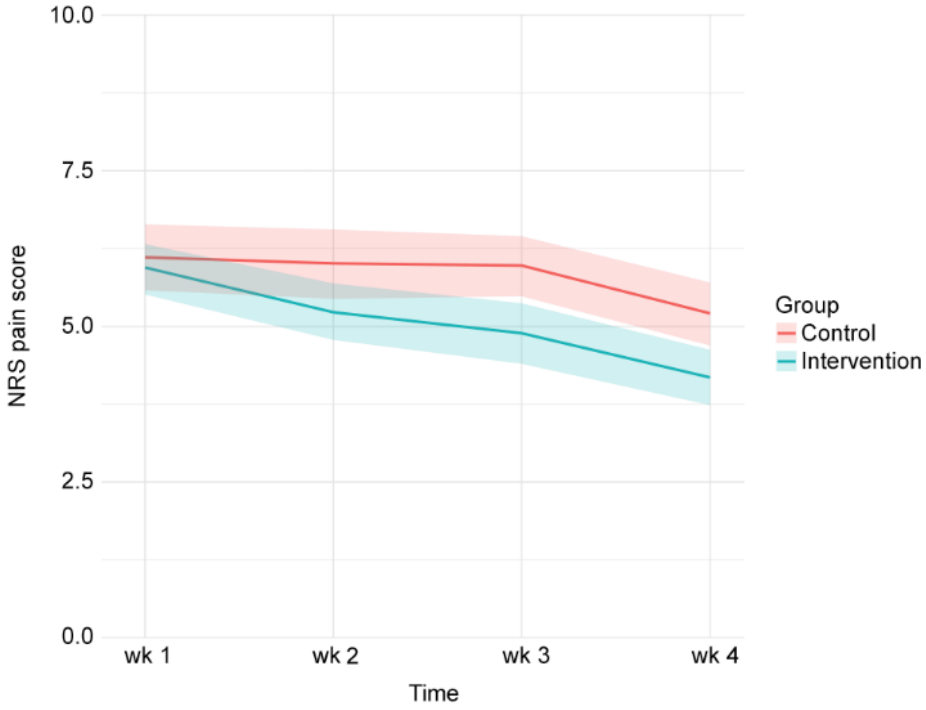


Table 4. Level of pain.

Wk	Pain at rest				Pain during activity				Pain during the night			
	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value
1	5.20	5.54	0.25 (–0.41 to 0.90)	.46	5.86	5.65	–0.20 (–0.82 to 0.42)	.52	5.94	6.11	0.08 (–0.58 to 0.74)	.81
2	4.40	5.38	0.96 (0.31 to 1.62)	.004	5.01	5.84	0.81 (0.19 to 1.43)	.01	5.23	6.01	0.80 (0.14 to 1.46)	.02
3	4.11	5.05	0.88 (0.21 to 1.54)	.001	4.56	5.16	0.60 (0.03 to 1.23)	.06	4.89	5.98	1.05 (0.37 to 1.72)	.002
4	3.45	4.59	1.07 (0.42 to 1.73)	.001	3.99	5.08	1.11 (0.48 to 1.73)	<.001	4.18	5.21	1.02 (0.36 to 1.68)	.003

Secondary Outcomes

Physical Functioning and Quality of Life

Four weeks after discharge, patients in the intervention group reported a significant, 14% decrease in functional limitations (Knee injury and Osteoarthritis Outcome Score short form [KOOS PS]) compared to the control group (P<.001) (Table 5). With respect to quality of life (EQ-5D), patients in the intervention group also reported a significant, 14% increase compared to the control group (P<.001).

Table 5. Physical functioning and quality of life.

	Baseline	Week 4
Physical functioning		
Intervention group, mean (SD)	45.91 (12.77)	37.61 (10.17)[a]
Control group, mean (SD)	43.83 (10.87)	43.08 (12.96)[a]
Quality of life		
Intervention group, mean (SD)	0.66 (0.16)	0.76 (0.16)[a]
Control group, mean (SD)	0.65 (0.24)	0.67 (0.25)[a]

[a]P<.001 (Intervention Group versus Control Group at week 4).

Performing Physiotherapy Exercises and Activities of Daily Self-Care

In both the intervention and control group, the patients' ability to perform their physiotherapy exercises and daily self-care activities during the first four weeks after discharge increased. Patients in the intervention group performed better on each of the dimensions from the second week onwards (Table 6).

Table 6. Ability to perform physiotherapy exercises and activities of daily self-care.

	Performing physiotherapy exercises				Performing daily self-care activities			
	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value
Week 1	6.54	6.52	0.01 (–0.53 to 0.51)	.98	6.73	6.30	–0.45 (–0.90 to 0.0)	.05
Week 2	7.28	6.53	–0.72 (–1.24 to –0.19)	.001	7.84	6.98	–0.83 (–1.28 to –0.37)	<.001
Week 3	7.35	6.93	–0.41 (–0.94 to 0.13)	.13	8.19	7.36	–0.79 (–1.26 to –0.33)	.001
Week 4	7.50	6.88	–0.60 (–1.12 to –0.08)	.03	8.32	7.64	–0.67 (–1.21 to –0.21)	.004

Satisfaction with Information and Patient-Perceived Involvement by the Hospital

In both the intervention and control group, patients' satisfaction with the provided information and patients' perceptions on how the hospital was involved in their recovery process decreased during the first four weeks after discharge. However, patients in the intervention group demonstrated a much smaller decrease over time (Table 7).

Table 7. Satisfaction with information and patient-perceived involvement by the hospital.

Week	Satisfaction with the Information				Patient-perceived involvement			
	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value
1	7.90	7.40	–0.51 (–1.08 to –1.80)	.07	7.52	6.87	–0.74 (–1.36 to 0.11)	.002
2	7.96	6.61	–1.38 (–1.95 to –0.81)	<.001	7.75	6.14	–1.64 (–2.27 to –1.0)	<.001
3	7.45	6.16	–1.29 (–1.87 to –0.71)	<.001	7.23	6.13	–1.13 (–1.78 to –0.48)	<.001
4	7.61	5.32	–2.28 (–2.85 to –1.71)	<.001	7.24	4.90	–2.35 (–2.99 to 1.79)	<.001

Health Care Consumption

Patients in the intervention group had, on average, 1.22 points of contact with the hospital, their general practitioner (GP), or home-care organization compared to 1.62 points of contact in the control group, which is a 33% difference (*P*=.014). Of all contacts by the intervention group, 36% did not lead to an action (consultation or referral), compared to 45% in the control group, a 25% difference (*P*=.14) (Table 8).

Table 8. Health care consumption.

	Hospital	GP[a]	Home care	Total	Average points of contact[b]
Intervention (n=105)	69	34	25	128	1.22
No action required	26	14	6	46	_[c]
Control (n=90)	59	58	28	145	1.62
No action required	23	30	12	65	–

[a]GP: general practitioner.

[b]The average number of contacts with the hospital, GP, or home-care organization (corrected for the time points patients participated in the measurement).

[c]Not applicable.

App Usage Data

In total, patients in the intervention group used the app 2418 times, which was an average of 26 times per patient. Most patients used a smartphone to access the information (75% vs 25% tablet use). The app was primarily used in the first 2 weeks, in which most of the information was offered (26/32 unique information items). Text-only information items related to pain,

wound care, and self-care activities were consulted most frequently. Information about the start of the third and fourth week and reminders for anticoagulant usage and participation in study questionnaires were the least consulted information items.

Over the course of the four-week intervention, an average of 25 videos were offered to each patient. In total, these videos were viewed 2950 times, which was an average of 36 views per patient. Video-enriched information items related to physiotherapy exercises, pain, and wound care were viewed most frequently. Videos related to wearing a Thrombo-Embolic Deterrent (TED) hose, reminders for usage of anticoagulation medication, and information on what to expect in the fourth week were the least frequently viewed videos.

Per Protocol Analysis

All results presented so far were analyzed using the intention-to-treat method. Analysis based on the per protocol method also resulted in the primary outcome (level of pain) being in favor of the intervention group, albeit somewhat more pronounced (Table 9).

Table 9. Per protocol analysis for level of pain.

Week	Pain at rest				Pain during activity				Pain during the night			
	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value	Intervention group, mean	Control group, mean	Model estimated difference (95% CI)	*P* value
1	4.98	5.54	0.54 (–0.15 to 1.23)	.13	5.71	5.65	–0.20 (–0.64 to 0.68)	.95	5.95	6.11	0.16 (–0.55 to 0.87)	.65
2	4.37	5.38	1.08 (0.38 to 1.77)	.002	4.94	5.84	0.99 (0.33 to 1.66)	.003	5.20	6.01	0.90 (0.19 to 1.61)	.01
3	4.05	5.05	1.01 (0.30 to 2.80)	.001	4.65	5.16	0.63 (0.04 to 1.30)	.06	4.91	5.98	1.09 (0.38 to 1.81)	.002
4	3.31	4.59	1.26 (0.56 to 1.96)	<.001	3.83	5.08	1.34 (0.68 to 2.00)	<.001	4.04	5.21	1.18 (0.47 to 1.89)	.001

Discussion

Primary Findings

The results of our study demonstrate the effectiveness of using an app to actively educate patients on a day-to-day basis in the first four weeks of their recovery after TKR. Regarding the primary outcome, patients in the intervention group experienced a lower level of pain while at rest, during activity and at night. Furthermore, the app had a positive effect on physical functioning, quality of life, performing physiotherapy exercises and activities of daily self-care, satisfaction with the information, how the hospital was involved in the recovery process, and health care consumption. Finally, the intervention resulted in a trend towards less health care consumption.

To our knowledge, our study is the first to assess the effectiveness of an app to improve TKR patients' self-management during the first four weeks of their recovery. In 2014, a Cochrane review on self-management programs for osteoarthritis showed that these programs (mainly including face-to-face and by-phone educational interventions) did not substantially improve self-management skills, pain, symptoms, physical functioning, or quality of life [29]. Furthermore, the authors concluded that new trials are unlikely to change the conclusions substantially, unless new models of self-management programs are introduced. In recent years, new self-management programs for patients (TKR included) have been introduced and have demonstrated the use and effectiveness of eHealth interventions, such as websites and portals [21]. An important difference between online interventions and app-based interventions like the one used in this trial is the ability to use push notifications. By using these notifications, information can be actively sent to patients at the time it is actually relevant to them instead of providing them with all the information all at once. Studies using apps in the TKR population have been performed but have focused on physiotherapy outcomes [28,30] and knowledge acquisition [31]. A 2018 multi-stakeholder analysis identified information needs during TKR treatment and suggested the use of technology to offer the right information at the right time [32]. As a result, patients would be allowed to better absorb the information and improve information recall and compliance. From a more clinical perspective, Filardo et al have suggested developing cointerventions to overcome kinesiophobia, the fear of physical activity due to pain. Kinesiophobia has significant impact upon patient recovery and final outcomes after TKR [33]. The app used in our study could be considered one of these cointerventions as it effectively lowered the level of pain in the intervention group.

With regards to economic benefits, using the app resulted in a decrease in health care consumption within the intervention group. This is in line with a 2017 review on the economic evidence for mobile health interventions, in which apps are reported to be an ideal platform for behavioral change in patients

because of their popularity, connectivity, and increased sophistication when compared to text message or telephone follow-up [34]. A decrease in health care consumption is of clinical relevance, since the minimized length of stay for TKR has proven to increase the burden of care on hospital staff [35].

A major strength of our study was the holistic approach to the early-phase recovery of TKR patients, combining insights from orthopedic surgeons, physiotherapists, and (specialized) nurses. In addition, historical data on TKR health care consumption was used to form an interactive timeline to guide patients during their recovery. Using push notifications to actively alert patients about newly available information resulted in an average use of the app of 26 times per patient. Additionally, the usage of short, one subject only videos has proven to be of added value, as patients watched many of the 25 available videos multiple times. This was especially the case with videos related to pain, wound care, and physiotherapy exercises, demonstrating patients' needs for this type of information.

A limitation of our study is the number of patients in the intervention group that did download and use the app. Out of the 114 possible app users, 93 patients downloaded the app (82%). This demonstrates the necessity of assessing a patient's digital health literacy and supporting them during the initial usage of interventions like these. Nevertheless, a per protocol analysis showed similar results. Another limitation could be the usage of self-developed questions which might introduce a risk of bias. To minimize this, questions were developed together with health care providers, were screened for readability by a specialized organization and were evaluated by several patients. Finally, we did not consider direct patient feedback when developing the content for the app but instead based it solely on the knowledge and experience of hospital staffs. In future research, patients should have a more prominent role to play in the development and evaluation of the app's (patient-specific) content, the desired timing of notifications, and the preferred mode of information to further optimize outcomes. In addition, future research could focus on the generalizability of interventions like these in other treatments, as well as their cost-effectiveness.

Conclusion

In conclusion, we found that, in comparison with standard patient education, the active education and coaching of patients on a day-to-day basis via the app in the four weeks after TKR resulted in a significant decrease in patients' levels of pain. Additionally, there was a significant improvement in patients' physical functioning, quality of life, their ability to perform physiotherapy exercises and activities of daily self-care, their satisfaction with the information, their perceived involvement by the hospital, and their health care consumption. Given the rising number of TKR patients and the increased emphasis on self-management, we suggest using an app with timely postoperative care education as a standard part of care.

Acknowledgments

The authors would like to thank all the participating hospitals, their medical professionals, research coordinators, and research assistants for their contribution to the success of this study. Furthermore, they would like to thank the team of Interactive Studios for creating the app and Sam Credgington for language editing. Finally, they would like to thank all patients who were willing to participate in the study.

Authors' Contributions

TT conceived the study and designed the trial, with the help of LJ, WW, RK and OLH. TT, WW, BZ, and BT supervised the data collection. TT managed the data. TT, LJ, WW and GH provided statistical analysis. OLH, WJM, DD, AP, and JWS provided the content used on the patients' timeline. TT drafted the manuscript. All authors contributed to its revision.

Conflicts of Interest

The principal investigator, Thomas Timmers, is one of the cofounders of Interactive Studios. Interactive Studios is the company that developed the app used in this study. Interactive Studios offered the app used in this study free of charge.

Multimedia Appendix 1
Overview of information offered to patients in the intervention group.
[PDF File (Adobe PDF File), 95 KB - mhealth_v7i10e15323_app1.pdf]

Multimedia Appendix 2
Self-developed questions for the study.
[PDF File (Adobe PDF File), 62 KB - mhealth_v7i10e15323_app2.pdf]

Multimedia Appendix 3
Consort eHealth checklist.
[PDF File (Adobe PDF File), 2369 KB - mhealth_v7i10e15323_app3.pdf]

References

1. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013 Sep;21(9):1145-1153 [FREE Full text] [doi: 10.1016/j.joca.2013.03.018] [Medline: 23973124]
2. Fingar K, Stocks C, Weiss A, Steiner C. Most Frequent Operating Room Procedures Performed in U.S. Hospitals. Rockville: Agency for Healthcare Research and Quality; 2003.
3. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007 Apr;89(4):780-785. [doi: 10.2106/JBJS.F.00222] [Medline: 17403800]
4. Kahlenberg CA, Nwachukwu BU, McLawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient Satisfaction After Total Knee Replacement: A Systematic Review. HSS J 2018 Jul;14(2):192-201 [FREE Full text] [doi: 10.1007/s11420-018-9614-8] [Medline: 29983663]
5. Husted H, Solgaard S, Hansen TB, Søballe K, Kehlet H. Care principles at four fast-track arthroplasty departments in Denmark. Dan Med Bull 2010 Jul;57(7):A4166. [Medline: 20591341]
6. Kehlet H, Søballe K. Fast-track hip and knee replacement--what are the issues? Acta Orthop 2010 Jun;81(3):271-272 [FREE Full text] [doi: 10.3109/17453674.2010.487237] [Medline: 20446827]
7. Kehlet H. Fast-track hip and knee arthroplasty. Lancet 2013 May 11;381(9878):1600-1602. [doi: 10.1016/S0140-6736(13)61003-X] [Medline: 23663938]
8. Vehmeijer SBW, Husted H, Kehlet H. Outpatient total hip and knee arthroplasty. Acta Orthop 2018 Apr;89(2):141-144 [FREE Full text] [doi: 10.1080/17453674.2017.1410958] [Medline: 29202644]
9. World HO. WHO Regional Office for South-East Asia. 2014. Self care for health URL: https://apps.who.int/iris/bitstream/handle/10665/205887/B5084.pdf?sequence=1&isAllowed=y [accessed 2019-09-26]
10. Chan E, Blyth FM, Cheow S, Fransen M. Postoperative pain following hospital discharge after knee replacement surgery: a patient survey. Pain Manag 2013 May;3(3):177-188. [doi: 10.2217/pmt.13.14] [Medline: 24654761]
11. Chan EY, Blyth FM, Nairn L, Fransen M. Acute postoperative pain following hospital discharge after total knee arthroplasty. Osteoarthritis Cartilage 2013 Sep;21(9):1257-1263 [FREE Full text] [doi: 10.1016/j.joca.2013.06.011] [Medline: 23973139]
12. McDonall J, de Steiger R, Reynolds J, Redley B, Livingston P, Botti M. Patient participation in postoperative care activities in patients undergoing total knee replacement surgery: Multimedia Intervention for Managing patient Experience (MIME). Study protocol for a cluster randomised crossover trial. BMC Musculoskelet Disord 2016 Jul 18;17:294 [FREE Full text] [doi: 10.1186/s12891-016-1133-5] [Medline: 27431239]
13. Zhu N, Xu P, Lei T, Sun T, Chan SW. Postoperative Pain Self-Management Behavior in Patients Who Underwent Total Knee or Hip Arthroplasty. AORN J 2017 Apr;105(4):355-364. [doi: 10.1016/j.aorn.2017.02.001] [Medline: 28336024]
14. Fredericks S, Guruge S, Sidani S, Wan T. Postoperative patient education: a systematic review. Clin Nurs Res 2010 May;19(2):144-164. [doi: 10.1177/1054773810365994] [Medline: 20435785]
15. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [doi: 10.1258/jrsm.96.5.219] [Medline: 12724430]
16. Specht K, Agerskov H, Kjaersgaard-Andersen P, Jester R, Pedersen BD. Patients' experiences during the first 12 weeks after discharge in fast-track hip and knee arthroplasty - a qualitative study. Int J Orthop Trauma Nurs 2018 Nov;31:13-19. [doi: 10.1016/j.ijotn.2018.08.002] [Medline: 30297138]
17. Szöts K, Pedersen PU, Hørdam B, Thomsen T, Konradsen H. Physical health problems experienced in the early postoperative recovery period following total knee replacement. Int J Orthop Trauma Nurs 2015 Feb;19(1):36-44. [doi: 10.1016/j.ijotn.2014.03.005] [Medline: 25787815]
18. Campbell R, Evans M, Tucker M, Quilty B, Dieppe P, Donovan JL. Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. J Epidemiol Community Health 2001 Feb;55(2):132-138 [FREE Full text] [doi: 10.1136/jech.55.2.132] [Medline: 11154253]
19. Güney-Deniz H, Irem Kınıklı G, Çağlar Ö, Atilla B, Yüksel İ. Does kinesiophobia affect the early functional outcomes following total knee arthroplasty? Physiother Theory Pract 2017 Jun;33(6):448-453. [doi: 10.1080/09593985.2017.1318988] [Medline: 28481125]
20. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for Patient Engagement: A Systematic Review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
21. van der Meij E, Anema JR, Otten RHJ, Huirne JAF, Schaafsma FG. The Effect of Perioperative E-Health Interventions on the Postoperative Course: A Systematic Review of Randomised and Non-Randomised Controlled Trials. PLoS One 2016;11(7):e0158612 [FREE Full text] [doi: 10.1371/journal.pone.0158612] [Medline: 27383239]
22. IQVIA. 2017 Nov 7. The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-09-26]
23. ITU. Global ICT developments. Geneva, Switzerland: International Telecommunication Union (ITU); 2018:2001.
24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]

26. Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008 May;16(5):542-550 [FREE Full text] [doi: 10.1016/j.joca.2007.12.014] [Medline: 18294869]
27. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
28. Hardt S, Schulz MRG, Pfitzner T, Wassilew G, Horstmann H, Liodakis E, et al. Improved early outcome after TKA through an app-based active muscle training programme-a randomized-controlled trial. Knee Surg Sports Traumatol Arthrosc 2018 Nov;26(11):3429-3437. [doi: 10.1007/s00167-018-4918-2] [Medline: 29589050]
29. Kroon FPB, van der Burg LRA, Buchbinder R, Osborne RH, Johnston RV, Pitt V. Self-management education programmes for osteoarthritis. Cochrane Database Syst Rev 2014 Jan 15(1):CD008963. [doi: 10.1002/14651858.CD008963.pub2] [Medline: 24425500]
30. Dallimore R, Asinas-Tan ML, Chan D, Hussain S, Willett C, Zainuldin R. A randomised, double-blinded clinical study on the efficacy of multimedia presentation using an iPad for patient education of postoperative hip surgery patients in a public hospital in Singapore. Singapore Med J 2017 Sep;58(9):562-568 [FREE Full text] [doi: 10.11622/smedj.2016084] [Medline: 27121921]
31. Timmers T, Janssen L, Pronk Y, van der Zwaard BC, Koëter S, van Oostveen D, et al. Assessing the Efficacy of an Educational Smartphone or Tablet App With Subdivided and Interactive Content to Increase Patients' Medical Knowledge: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e10742 [FREE Full text] [doi: 10.2196/10742] [Medline: 30578185]
32. van Kasteren Y, Freyne J, Hussain MS. Total Knee Replacement and the Effect of Technology on Cocreation for Improved Outcomes and Delivery: Qualitative Multi-Stakeholder Study. J Med Internet Res 2018 Mar 20;20(3):e95 [FREE Full text] [doi: 10.2196/jmir.7541] [Medline: 29559424]
33. Filardo G, Merli G, Roffi A, Marcacci T, Berti Ceroni F, Raboni D, et al. Kinesiophobia and depression affect total knee arthroplasty outcome in a multivariate analysis of psychological and physical factors on 200 patients. Knee Surg Sports Traumatol Arthrosc 2017 Nov;25(11):3417-3423. [doi: 10.1007/s00167-016-4201-3] [Medline: 27329175]
34. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
35. Shah RP, Karas V, Berger RA. Rapid Discharge and Outpatient Total Joint Arthroplasty Introduce a Burden of Care to the Surgeon. J Arthroplasty 2019 Jul;34(7):1307-1311. [doi: 10.1016/j.arth.2019.03.052] [Medline: 31031153]

Abbreviations

ANOVA: analysis of variance
CONSORT: Consolidated Standards of Reporting Trials
eHealth: electronic health
GP: general practitioner
mHealth: mobile health
NRS: Numeric Rating Scale
TED: Thrombo-Embolic Deterrent
TKR: total knee replacement

Edited by G Eysenbach; submitted 01.07.19; peer-reviewed by JB Park, J Floch, H Mehdizadeh, A Graham, R Grainger, K Fitzner, T Wyatt, JR Bautista; comments to author 24.07.19; revised version received 24.08.19; accepted 23.09.19; published 21.10.19.

Please cite as:
Timmers T, Janssen L, van der Weegen W, Das D, Marijnissen WJ, Hannink G, van der Zwaard BC, Plat A, Thomassen B, Swen JW, Kool RB, Lambers Heerspink FO
The Effect of an App for Day-to-Day Postoperative Care Education on Patients With Total Knee Replacement: Randomized Controlled Trial
JMIR Mhealth Uhealth 2019;7(10):e15323
URL: https://mhealth.jmir.org/2019/10/e15323
doi:10.2196/15323
PMID:31638594

Original Paper

The Influence of Physician Information on Patients' Choice of Physician in mHealth Services Using China's Chunyu Doctor App: Eye-Tracking and Questionnaire Study

Wei Shan[1], PhD; Ying Wang[1], BSc; Jing Luan[2], PhD; Pengfei Tang[1], BSc

[1]School of Economics and Management, Beihang University, Beijing, China

[2]School of Economics and Management, Beijing Jiaotong University, Beijing, China

Corresponding Author:
Jing Luan, PhD
School of Economics and Management
Beijing Jiaotong University
3 Shangyuancun, Haidian District
Beijing, 100044
China
Phone: 86 18146510686
Email: jingluan@bjtu.edu.cn

Abstract

Background: Mobile health (mHealth) is becoming more popular as a way of sharing medical information. For the patient, it saves time, reduces the need for travel, reduces the cost of searching for information, and brings medical services "to your fingertips." However, it also brings information overload and makes the patient's choice of physician more difficult.

Objective: This study aimed to identify the types of physician information that play a key role in patients' choice of physician and to explore the mechanism by which this information contributes to this choice.

Methods: Based on the stimulus-organism-response (SOR) model and online trust theory, we proposed a research model to explain the influence of physician information on patients' choice of physician. The model was based on cognitive trust and affective trust and considered the moderating role of patient expertise. Study 1 was an eye-tracking experiment (n=42) to identify key factors affecting patients' choice of physician. Study 2 was a questionnaire study (n=272); Partial Least Squares Structural Equation Modeling was used to validate the research model.

Results: The results of Study 1 revealed that seven types of physician information played a key role in patients' choice of physician. The results of Study 2 revealed that (1) physicians' profile photo information affected patients' choice of physician by positively influencing affective trust (P<.001); (2) physicians' nonprofile photo information affected patients' choice of physician by positively influencing cognitive trust (P<.001); (3) patient-generated information affected patients' choice of physician by positively affecting cognitive trust (P<.001) and affective trust (P<.001), and patient expertise played a positive moderating role on both (P=.04 and P=.01, respectively); and (4) cognitive trust and affective trust both positively affected patients' choice of physician, with affective trust playing a more significant role (P<.001 and P<.001, respectively).

Conclusions: Seven types of physician information were mainly used by patients when choosing physicians offering mHealth services; trust played an important role in this choice. In addition, the level of patient expertise was an important variable in moderating the influence of physician information and patients' trust. This paper supports the theoretical basis of information selection and processing by patients. These findings can help guide app developers in the construction of medical apps and in the management of physician information in order to facilitate patients' choice of physician.

(JMIR Mhealth Uhealth 2019;7(10):e15544) doi:10.2196/15544

KEYWORDS

mHealth; physician information; choice; trust

Introduction

Background

Information asymmetry between patients and physicians is one of the main causes of tension in the physician-patient relationship within traditional medical services [1]. Due to the lack of professional knowledge, patients are in a relatively disadvantaged position; therefore, promoting physician-patient interaction and reducing information asymmetry has been one of the focus areas of the medical industry. The term mobile health (mHealth) refers to the provision of medical information and services through mobile communication technologies and mobile devices, such as mobile phones [2]. In recent years, with the rapid development of Internet technology, the influence of mHealth in our lives has grown [3-5]. mHealth can improve the efficiency of medical services and reduce their cost through remote medical monitoring and consultation [6,7]. This new pattern of physician-patient interaction and sharing of physician information can help solve current problems in the field of medicine, such as medical inefficiency and physician-patient contradictions.

Previous research on mHealth mainly focused on its development trend [8-10]; the function, development, and design of mHealth apps [11-13]; and users' intentions to adopt and use mHealth services [14-16]. However, in the field of mHealth user behavior, there is still a lack of in-depth research on patients' choice of physician and its influencing factors. In fact, patients' choice of physicians who offer mHealth is different from the choice they make within the traditional route. First, mHealth provides information about physicians far beyond the traditional medical model. On one hand, such physician information can change patients' passive positions when accessing medical services and can improve the physician-patient relationship, further affecting their choice of physician [17-19]. On the other hand, it can also lead to information overload to some extent, which has side effects on patients; it can cause patients to feel confused when making medical choices, which could reduce their efficiency [20]. Therefore, it is important to find out what information plays a major role in patients' choice of physician and to conduct effective information management to help patients. In addition, unlike the traditional physician-patient relationship, in the mHealth environment patients learn about physicians through mHealth platforms; they connect with each other through the network rather than via direct contact. Researchers have shown that health care is a service that requires a high amount of trust on the part of the patient [21]; in addition, trust is important for the success of online health services [22]. Therefore, patients' trust in physicians is important for the relatively unfamiliar physician-patient relationship in mHealth. Based on the above considerations, this study focused on two important issues in mHealth: (1) the impact of physician information on patients' choice of physician and (2) the trust mechanism between physician information and patients' choice of physician.

To address these issues, this study focused on the Chunyu Doctor app, which is the largest mobile physician-patient communication platform in China and has attracted more than 500,000 physicians in public hospitals. Based on the stimulus-organism-response (SOR) model and online trust theory, this study explored how physician information affects patient trust in mHealth, thereby further affecting patients' choice of physician. The role of patient expertise in the relationship between physician information and trust was also considered. Since eye-tracking methods have been used to record user behavior [23], in this study we attempted to use eye tracking to investigate the key factors that influence patients' choice of physician. Further, a questionnaire was used to decipher the internal mechanism of this influence. This study provides a certain theoretical basis for the research of mHealth information services and provides practical guidance for the construction of mHealth apps and their information management methods.

Theoretical Foundation

Online Trust

Online trust theory is based upon many studies that have classified online trust as consumer trust in e-commerce [24,25]. When consumers feel that information is insufficient or asymmetric, trust can be used as a means to alleviate the asymmetry of perceived information, prompting them to make purchasing decisions [26,27]. McAllister divided trust into cognitive trust and affective trust [28]. Cognitive trust is mainly regarded as the consumer's rational expectation of the ability and credibility of the trusted party and is mostly judged by the objective characteristic evidence of the trusted party's personal behavior and reputation. Affective trust is regarded as a trust attitude, which reflects the consumer's feelings and self-consciousness toward the trusted party [29-31].

Choice in mHealth can also be regarded as a special online shopping behavior. In mHealth service, if the physician is regarded as the *commodity*, then the patient is equivalent to the *consumer*. Therefore, the study of the patient's choice in mHealth is actually the study of the consumption behavior of this particular consumer group. Asymmetry of information results in trust playing an important role in the physician-patient relationship. At present, physician-patient trust is generally defined from the patient's point of view; it is defined as the patient's trust in the physician's ability to diagnose and treat them and trust that the physician will put their interests first [32]. Therefore, patient trust plays a very important role in constructing the patient choice model. In this study, trust is also divided into cognitive trust and affective trust. In addition, according to the study by Calefato et al [33], the patient's trust in the physician's abilities (ie, physician's professional skills, knowledge, and competence) is defined as the patient's cognitive trust; the patient's trust in the physician's benevolence (ie, physician's politeness, attitude, and willingness to help) is defined as the patient's affective trust.

Stimulus-Organism-Response Model

The SOR model, proposed by Mehrabian and Russell in 1974, illustrates that the external environment influences the individual's attitude or behavior by influencing the individual's mental state. *Stimulus* refers to the external environmental factors received by the individual; *organism* refers to the internal mental state of the individual; and *response* refers to the attitude

and behavior of the individual [34]. The model has been widely used in consumer behavior analysis. According to the model, a consumer (ie, *organism*) will have a conscious or unconscious psychological reaction after receiving an external *stimulus*, which involves both cognitive and affective aspects, and will then have an internal or external behavioral *response* to this stimulus [35-37]. In this study, physician information in the mHealth app, Chunyu Doctor, was considered an external *stimulus* that caused patients (ie, *organisms*) to generate trust, which in turn allowed them to make their choice of physicians (ie, *response*).

Research Model and Hypotheses

Physician Information and Trust

Physicians are the direct providers of medical services. In the field of mHealth, physician information is one of the most important considerations for patients when choosing medical services. Patients can get detailed information about a physician when browsing the physician's home page. This study extracted physician information from two major information sources: physicians and patients. Accordingly, three kinds of information were identified: physicians' profile photo information, physicians' nonprofile photo information, and patient-generated information.

In the past physician-patient relationship studies, information from physicians' profile photos was often neglected, but the face is an important source of information. Sometimes a photo can promote trust between people [38]. This face-based assessment is very subjective [39], with emotion often generating trust in others, while physicians' nonprofile photo information is different. For example, physicians in high-level hospitals often have more advanced medical skills and richer medical experience. As well, a physician's title is a reflection of their professional skills, medical experience, and work performance. These personal traits often lead patients to trust physicians at the cognitive level.

In online trust theory, user-generated information can enhance the trust of online sellers, reduce the risk perceived by buyers, and promote transaction behavior [40]. User-generated information in mHealth corresponds to patient-generated information. The main forms of this information are comments, feedback, and scoring, which can reflect online reputation [41]. Patient-generated information is usually the first-hand experience of patients with online consultations. Eysenbach proposed that, in the field of health care, experience-based credibility can be seen as an additional dimension of source credibility [42]. In that sense, similarity between experiences will enhance patients' perceptions of credibility. In addition, some studies have suggested that patient-generated information can reveal physicians' online behavior; reduce information asymmetry between physicians and patients [43]; prevent physicians from exaggerating and misinterpreting their professional ability, service level, and treatment effect; and help patients understand physicians' medical expertise and service attitude [44]. In conclusion, patient-generated information will affect patients' trust in physicians at the cognitive and affective levels, thus affecting their behavior in choosing physicians. Therefore, the following hypotheses are put forward:

Hypothesis 1a (H1a): A physician's nonprofile photo information will positively affect the patient's cognitive trust.

Hypothesis 1b (H1b): A physician's profile photo information will positively affect the patient's affective trust.

Hypothesis 1c (H1c): Patient-generated information will positively affect the patient's cognitive trust.

Hypothesis 1d (H1d): Patient-generated information will positively affect the patient's affective trust.

Patient Expertise, Physician Information, and Trust

Consumer expertise is a key factor influencing a consumer's choice; it is the knowledge and experience, of which a consumer is subjectively aware, regarding a particular product or service [45]. High-expertise consumers are willing to take the initiative to look for information related to products and to make decisions through careful thinking because of their rich experience. Low-expertise consumers are more passive in searching for information, preferring to rely on marginal information, and are less willing to spend time thinking about new information.

In consumer behavior research, consumer expertise plays a significant role in explaining consumer decisions and responses to products [46,47]. In this study, patient expertise refers to the patient's knowledge or experience of mHealth services. Since patient-generated information is unique information in mHealth compared with traditional medical services, we believe that patient expertise will play a moderating role between physician information and trust. That is, the higher the expertise of patients, the more they will rely on patient-generated information in choosing a physician; however, low-expertise patients will rely more on the physician's profile photo and nonprofile photo information. Hence, we hypothesize the following:

Hypothesis 2a (H2a): Patient expertise will play a negative role in the influence of a physician's nonprofile photo information on cognitive trust.

Hypothesis 2b (H2b): Patient expertise will play a negative role in the influence of a physician's profile photo information on affective trust.

Hypothesis 2c (H2c): Patient expertise will play a positive role in the impact of patient-generated information on cognitive trust.

Hypothesis 2d (H2d): Patient expertise will play a positive role in the impact of patient-generated information on affective trust.

Trust and Choice of Physician

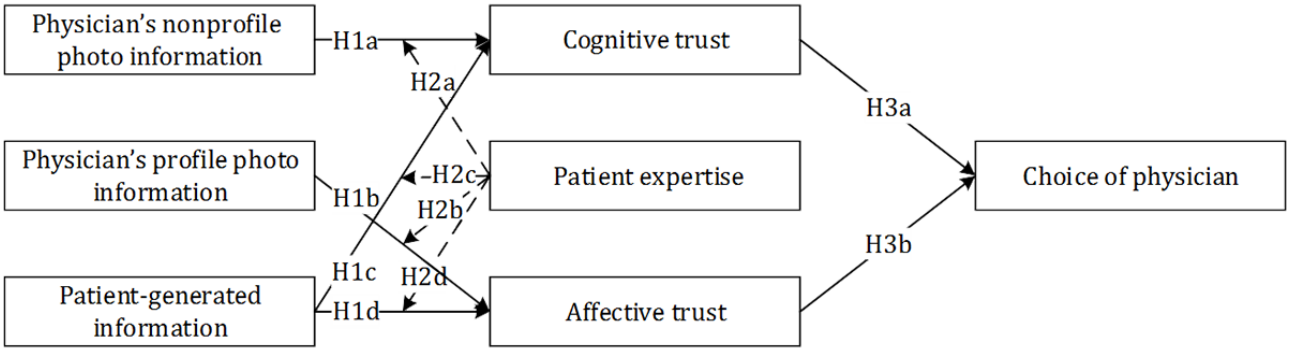
Corritore and Wiedenbeck [22] proposed that in online medical care, physician-patient trust is one of the most important factors to ensure the success of online health services. Gefen et al [48] believed that consumers' intentions to use certain goods or services mainly depended on consumers' cognitive trust and affective trust in external factors. Therefore, in this study, patients' trust is considered to influence patients' choice of physician in mHealth. Hence, we hypothesize the following:

Hypothesis 3a (H3a): The patient's cognitive trust in a physician will positively influence his or her choice of physician.

Hypothesis 3b (H3b): The patient's affective trust in a physician will positively influence his or her choice of physician.

Based on our hypotheses, this study illustrates a research model that explains the influence of physician information on the patients' choice of physician in mHealth, as shown in Figure 1.

Figure 1. Research model explaining the influence of physician information on the patients' choice of physician in mHealth. H1a: hypothesis 1a; H1b: hypothesis 1b; H1c: hypothesis 1c; H1d: hypothesis 1d; H2a: hypothesis 2a; H2b: hypothesis 2b; H2c: hypothesis 2c; H2d: hypothesis 2d; H3a: hypothesis 3a; H3b: hypothesis 3b.



Methods

Overview

Two studies were conducted. Study 1 aimed to identify the key physician information that had an impact on patients' choice of physician in mHealth through an eye-tracking experiment. Study 2 used questionnaires to explore what influence physician information, determined in Study 1, had on patients' choice of physician.

Study 1: Investigating Physician Information

Research Design

The purpose of this study was to investigate the impact that physician information had on patients' choice of physician in an mHealth platform and to identify the information that played a key role in this choice. Research has shown that people's ability to process information within their short-term memory is limited [49]; therefore, even though the information provided in the mHealth app, Chunyu Doctor, is comprehensive, a patient's choice of physician is based only on a limited amount of useful information. In previous online medical studies, scholars have often used their own experiences and related theories to select the information they think is more useful, built regression models by crawling data to uncover the relationship between the physician information and the number of physicians' consultations, and judged the impact of physician information on patients' choice of physician [50-52]. This method is subjective and it is easy to overlook some potentially critical information. Humans obtain much external information through their vision and the images to which their attention is focused [53]. Through eye-tracking experiments, researchers can capture the information upon which people focus their attention more objectively and accurately. Therefore, in this study, eye-tracking experiments were used to screen physician information that had a key impact on patients' choice of physician.

The Chunyu Doctor app contains 12 different types of physician information; for the universality of the experimental results, we did not consider the physicians' departments or specific communication between physicians and patients. The types of information we were concerned with were divided into three categories: physicians' profile photo information, physicians' nonprofile photo information, and patient-generated information. This information is shown in Table 1.

In this study, the 12 types of physician information (see Table 1) were classified into high-level and low-level information. Hospital, title, and educational background levels were classified by the government (eg, hospitals are classified into Class A tertiary hospitals and general hospitals); because of this, other information levels, except physicians' profile photo information, were set with reference to the Chunyu Doctor app and all of them were pretested using a questionnaire. The questionnaire showed that the various information levels were significantly different; the specific classification is shown in Multimedia Appendix 1.

Table 1. Classification of physician information.

Information category	Types of information
Physicians' profile photo information	Profile photo
Physicians' nonprofile photo information	Hospital, title, educational background, academic research results, topic, fees, and peer evaluations
Patient-generated information	Consultation numbers, favorability rate, satisfaction, and gratitude expressed

Regarding physicians' profile photo information, based on a study by Ert et al [54], two dimensions were used to distinguish the information level: attractiveness and trustworthiness. A total of 60 participants—28 males (47%) and 32 females (53%)—rated 40 physician profile photos based on their perceived attractiveness and trustworthiness. We used photos of physicians from the official website of Peking Union Medical College Hospital; 20 males and 20 females, aged 30-40 years old, were selected. In the photos, all physicians wore glasses, white coats, and light sweaters or shirts; none had beards, accessories, or visible makeup. All photos were front-facing portraits of physicians smiling. Each participant had to look at all 40 physician profile photos. For each physician's profile photo, participants were asked to answer the following questions: "Do you think the physician is attractive?" and "Do you think the physician is trustworthy?" These were asked in order to estimate the perceived levels of attractiveness and trustworthiness of the profile photos. The questions were scored on a 7-item Likert scale, ranging from 1 (*totally unattractive/untrustworthy*) to 7 (*very attractive/trustworthy*); the order in which the photos appeared to participants was random. Eight profile photos were selected based on the photos' average attractiveness and trustworthiness levels; four photos had the highest and four photos had the lowest average attractiveness and trustworthiness levels, respectively. There was a significant difference between high-level and low-level profile photos (F_1=94.58, P<.001); however, there was no significant difference between the four high-level profile photos (F_3=0.622, P=.60) and the four low-level profile photos (F=0.292, P=.83). In addition, there was no significant difference between male and female physicians (F_1=2.547, P=.11).

In this study, 12 types of physician information were grouped (see Table 1) and divided into three major categories. Each category was classified according to the rated levels (eg, high-level patient-generated information contained high consultation numbers, high praise rates, high satisfaction, and high gratitude expressed; low-level patient-generated information contained low consultation numbers, low praise rates, low satisfaction, and low gratitude expressed). This was used to generate eight mHealth physician home pages. In summary, the eye-tracking experiment was a 2 (physicians' profile photo information: high or low level) × 2 (physicians' nonprofile photo information: high or low level) × 2 (patient-generated information: high or low level) within-group experiment, in which eye-tracking data and questionnaire data were collected.

Procedure

A total of 42 undergraduate and postgraduate students from Beihang University participated in the experiment; participants were publicly recruited via the Chinese social media platform WeChat. In order to avoid gender bias, 21 men (50%) and 21 women (50%), aged 20-25 years old, were recruited. All participants had normal visual acuity or corrected visual acuity, with astigmatism below 200 degrees, and had online shopping experience. Before beginning the experiment, each participant signed an informed consent letter and registered their personal information. After the experiment, participants were compensated with CNY ¥30.

The eye-tracking equipment used in this study was the Tobii T120 eye tracker (Tobii Technology). The eye tracker was integrated into a 17-inch, thin-film-transistor display with a resolution of 1280 × 1024 and a sampling rate of 120 Hz. The instrument has a built-in eye-tracking server; the participants did not need to wear any equipment but did need to watch the experimental material displayed on the monitor. While the participants looked at the screen, the eye tracker automatically recorded eye-tracking data of the participants. Eye-tracking data can be analyzed and extracted using the Tobii Studio software associated with the eye tracker.

In order to simulate the mHealth treatment situation as realistically as possible, the experimental interface was based on the physician's home page on the Chunyu Doctor app. In order to avoid interference from other information, only the aforementioned 12 types of physician information were included. Each physician's home page was the same size and, as much as possible, the display format of the various types of information was consistent with the original app. The experiment began with an introduction:

> *We assume that you or a family member has a chronic disease, such as a sports injury, psychological stress, chronic disease, cold, fever, physical discomfort, etc, and you want to consult with your physician about the relevant information and treatment of the disease. In this experiment, you will be shown information about eight different physicians. These physicians are on the Chunyu Doctor mHealth app and work in a department related to the disease about which you want a consultation. After reading information about each physician, you will need to rate the statement "I choose this physician." The options range from 1 (strongly disagree) to 7 (strongly agree). Please complete the questionnaire according to your subjective feelings.*

During the introduction to the experiment, two unfamiliar types of information were explained, namely, gratitude expressed (ie, offering extra thank-you tips) and physician's topic (ie, number of physician's articles published). Participants were able to view the experimental stimuli (ie, the Chunyu Doctor app) by clicking the left mouse button on the screen. In order to ensure that the participants were able to fully observe the information about the eight physicians, each physician's home page was fixed for 1 minute. The eye-tracking experiment was complete after participants browsed the home pages of the eight physicians. Participants were then asked to complete a questionnaire; participants needed to rate the usefulness of the 12 types of information by rating the following statement for each on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*): "I think this information is very useful for me in helping me choose this physician." At this point, the experiment has been completed.

Study 2: The Mechanism of Influence of Physician Information on Patients' Choice

Study 2 was based on the physician information identified from Study 1 that played a key role in patients' choice. In this part of the study, we investigated how physician information affected patients' choice of physician. Questionnaires were used to collect sample data through a 2 (physician's profile photo information: high or low level) × 2 (physician's nonprofile photo information: high or low level) × 2 (patient-generated information: high or low level), between-group experiment to verify our research model. We combined the three categories of physician information to generate home pages for each of the eight physicians. Only the information identified from Study 1 that played a key role in patients' choice appeared on the home pages in Study 2. However, because fees are often related to physicians' hospitals and titles, the impact of fees on trust was not considered in this study. Each participant only saw one of the eight physician home pages, which was chosen at random. The questionnaire was divided into three parts: (1) collection of demographic information (eg, age, gender, educational level, online shopping experience, and frequency of use of mHealth apps); (2) measurement of the moderator variable, patient expertise; and (3) participants' responses regarding cognitive trust, affective trust, and physician choice after observing each physician's home page. Each variable was measured using a 7-item Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The design of the questionnaire drew on the relevant literature; appropriate adjustments were made according to the content of this study to ensure the validity and reliability of the constructed measures, as shown in Table 2. Questionnaires were distributed through the Questionnaire Star platform (China Wise Talent Information Technology Co), a professional online questionnaire survey service in China.

Table 2. Constructs and corresponding items.

Construct	Items
Cognitive trust (CT) [55]	• CT1: This physician was competent and effective in meeting my needs. • CT2: This physician was capable and proficient. • CT3: This physician was very knowledgeable in his or her medical field.
Affective trust (AT) [33,55]	• AT1: This physician would act in my best interest. • AT2: If I required help, this physician would do his or her best to help me. • AT3: I think this physician is friendly and approachable.
Choice of physician (CP) [56]	• CP1: I would be willing to choose this physician. • CP2: I would be willing to recommend this physician to others. • CP3: I have positive things to say about this physician.
Patient expertise (PE) [46]	• PE1: I am knowledgeable about mHealth services. • PE2: I learn well about mHealth services. • PE3: I have rich experience in mHealth services.

Results

Study 1: Investigation of Physician Information

The area of interest (AOI) in eye-tracking experiments is the gaze area to which researchers pay attention and represents the stimulus information in which participants are interested. AOI is the basic unit of analysis in eye-tracking experiment results. According to the classification of physician information as shown in Figure 2, the interface of each physician's home page was divided into 12 corresponding AOIs: profile photo, hospital, title, consultation numbers, favorability rate, peer evaluation, fees, educational background, academic research results, topic, satisfaction, and gratitude expressed. Based on previous literature, the average duration of visual fixation for each AOI was selected as the eye-tracking analysis index for this experiment. The average duration of fixation is the average time in seconds of fixation per unit area in the AOI. The longer the average duration of fixation, the more interested the participants were in the test material or the more difficult it was for them to extract information [57]. In order to identify the information that participants were more interested in rather than information that was difficult to extract based on the eye-tracking index, it was necessary to combine the results of the questionnaire. John Miller, an American psychologist, has accurately measured the capacity of short-term memory; this capacity for a normal adult is 7 ± 2 items at a time, with memory ability decreasing with more than seven items [49]. Therefore, this study used the average ranking of eye-tracking data (ie, average duration of fixation) and questionnaire data (ie, ranking of the usefulness of information) to ultimately select seven aspects of information that had a key impact on patients' choice of physician: favorability rate, consultation numbers, title, hospital, satisfaction, profile photo, and fees. The specific data are shown in Table 3.

As the results show in Table 3, we observed an interesting finding. People often subjectively believe that patient-generated information (ie, favorability rate, satisfaction, consultation numbers, and gratitude expressed) is more useful to them in choosing a physician than a physician's profile photo. This may also be the reason why physicians' profile photo information has not been taken into account in previous studies. However, eye-tracking data revealed contradictory results; according to this data, people payed much more attention to the profile photos. It can be seen that a physician's profile photo is also a very important factor in patients' choice of physician.

Figure 2. Area of interest division map of physician home pages.

Table 3. Ranking of 12 types of physician information.

Item #	Physician information	Duration of fixation (seconds), mean (SD)	Rank for duration of fixation	Information usefulness[a]	Rank for information usefulness	Ranking average
1	Favorability rate	8.14 (3.45)	2	5.88	1	1.5
2	Consultation numbers	7.22 (2.39)	4	5.74	3	3.5
3	Title	8.30 (4.55)	1	4.89	6	3.5
4	Hospital	6.74 (5.68)	5	5.08	5	5.0
5	Satisfaction	4.71 (2.31)	9	5.85	2	5.5
6	Profile photo	7.30 (3.96)	3	4.22	9	6.0
7	Fees	5.80 (1.72)	6	4.76	7	6.5
8	Gratitude expressed	4.50 (2.58)	10	5.58	4	7.0
9	Educational background	4.77 (1.28)	8	4.39	8	8.0
10	Peer evaluations	5.03 (2.16)	7	3.60	12	9.5
11	Academic research results	4.09 (1.25)	11	4.22	10	10.5
12	Topic	3.84 (1.55)	12	3.88	11	11.5

[a]Participants rated the usefulness of the 12 types of information by rating the following statement for each on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*): “I think this information is very useful for me in helping me choose this physician.”

Study 2: The Mechanism of Influence of Physician Information on Patients' Choice of Physician

Sample Characteristics

A total of 272 questionnaires were recovered, of which 254 (93.4%) were valid. The sample distribution characteristics are shown in Multimedia Appendix 2. Among the participants, 44.5% (113/254) were male and 55.5% (141/254) were female. A total of 87.4% (222/254) of the participants were under 30 years old and all of them had online shopping experience. In line with the fact that young people are more receptive to new things, mobile device users tend to be younger and older people mostly use mHealth apps via their children. A total of 42.1% (107/254) had never used mHealth services while 57.9% (147/254) had used them; 17.7% (45/254) had used them one or more times a week. These findings could help explain the patient expertise in mHealth of the study participants. Based on the above analysis, the study sampling was considered reasonable.

Reliability and Validity

The reliability and validity of the measurement instruments were tested [58,59], as shown in Tables 4 and 5. SPSS, version 22.0 (IBM Corp), was used to analyze the reliability of the data and to measure the Cronbach alpha values of each construct. The principal component analysis showed that the Kaiser-Meyer-Olkin value of the sample was 0.904, which indicated that it was suitable for factor analysis. SmartPLS 3.0 (SmartPLS GmbH) was used to carry out confirmatory factor experiments [60] to analyze the validity of the data, including convergence validity and discriminant validity. Table 4 shows that the Cronbach alpha values of all constructs were greater than .8 and that the composite reliability values were also greater than 0.8, which shows that the measured constructs had better reliability. The factor loads of all variables were greater than 0.7. The average variance extraction (AVE) values were greater than 0.5, which indicated that the constructs in this study had good convergence validity [61].

For the test of discriminant validity of the measurement scale, we can compare the square root of each factor's AVE and its correlation coefficients with other factors [61]. The results of the discriminant validity measurement are shown in Table 5, in which the enlarged diagonal values are the square root of each factor's AVE. Table 5 shows that the square root of each factor's AVE is greater than the corresponding correlation coefficients, which indicates that the discriminant validity between the constructs in this study was quite good.

Table 4. Construct reliability and convergence validity.

Construct amd items	Factor load	Cronbach alpha	Composite reliability	Average variance extraction
Cognitive trust (CT)		.950	0.968	0.909
CT1: This physician was competent and effective in meeting my needs.	0.938			
CT2: This physician was capable and proficient.	0.967			
CT3: This physician was very knowledgeable in his or her medical field.	0.955			
Affective trust (AT)		.903	0.940	0.838
AT1: This physician would act in my best interest.	0.885			
AT2: If I required help, this physician would do his or her best to help me.	0.937			
AT3: I think this physician is friendly and approachable.	0.924			
Choice of physician (CP)		.899	0.937	0.832
CP1: I would be willing to choose this physician.	0.932			
CP2: I would be willing to recommend this physician to others.	0.932			
CP3: I have positive things to say about this physician.	0.872			
Patient expertise (PE)		.906	0.941	0.841
PE1: I am knowledgeable about mHealth services.	0.916			
PE2: I learn well about mHealth services.	0.921			
PE3: I have rich experience in mHealth services.	0.914			

Table 5. Discriminant validity analysis.

Construct	Cognitive trust	Affective trust	Choice of physician	Patient expertise
Cognitive trust	0.953			
Affective trust	0.748	0.916		
Choice of physician	0.812	0.829	0.912	
Patient expertise	0.155	0.242	0.228	0.917

Analysis of Research Model

Partial Least Squares Structural Equation Modeling (PLS-SEM) has advantages over Covariance-Based Structural Equation Modeling (CB-SEM) or general linear regression modeling in processing sample data that do not obey normal distribution—physician information is a class variable—involving multiple constructs and small sample size prediction experiments [62]. Therefore, we used SmartPLS 3.0, an analysis software commonly used in partial least squares path modeling, to construct the whole structural equation; we used bootstrapping with 5000 resamples to assess the statistical significance of path coefficients [63].

The results show that 77.1% of the variation in patient choice could be explained by the model (R^2=0.771). Table 6 shows that most of the hypotheses in this model were supported. Physicians' nonprofile photo information had a significant positive impact on cognitive trust. Physicians' profile photo information had a significant positive impact on affective trust, which supports H1a and H1b. Patient-generated information had a significant positive impact on cognitive trust and affective trust, which supports H1c and H1d. Cognitive trust and affective trust had significant positive effects on patients' choice of physician, which supports H3a and H3b. However, the hypotheses regarding the moderating effect of patient expertise were only partially supported. Patient expertise had significant moderating effects on the influence of patient-generated information, both on cognitive trust and affective trust. That is, when patient expertise was higher, the level of information generated by patients was higher and they experienced more cognitive and affective trust; thus, H2c and H2d were supported. However, patient expertise had no significant moderating effect on the influence of physicians' nonprofile photo information on cognitive trust nor on physicians' profile photo information on affective trust. Therefore, H2a and H2b were not supported.

To sum up, this study showed that in mHealth, physician information affected patients' choice of a physician through trust. In addition, it should be noted that patients with different levels of expertise will have different experiences. The experimental results showed that patient expertise had no significant influence on the relationship between physicians' profile photo information and affective trust nor between physicians' nonprofile photo information and cognitive trust. It may be because physicians' profile photo information and physicians' nonprofile photo information are widely known types of information that exist in online medical treatment; no matter how much knowledge and experience patients have regarding mHealth care, it will not change the impact that these two types of information have on patients' choice of physician. However, patient expertise had a significant impact on the relationship between patient-generated information and cognitive trust and on the relationship between patient-generated information and affective trust. The interaction diagrams of the moderating effects are shown in Figures 3 and 4.

Table 6. Results of hypothesis testing.

Hypothesis	Path	Path coefficient	*t* test	*P* value	Supported?
H1a: A physician's nonprofile photo information will positively affect the patient's cognitive trust.	PNI[a]→CT[b]	0.258	4.558	<.001	Yes
H1b: A physician's profile photo information will positively affect the patient's affective trust.	PPI[c]→AT[d]	0.174	3.197	.001	Yes
H1c: Patient-generated information will positively affect the patient's cognitive trust.	PGI[e]→CT	0.218	3.764	<.001	Yes
H1d: Patient-generated information will positively affect the patient's affective trust.	PGI→AT	0.217	3.759	<.001	Yes
H2a: Patient expertise will play a negative role in the influence of a physician's nonprofile photo information on cognitive trust.	PNI × PE[f]→CT	0.097	0.933	.17	No
H2b: Patient expertise will play a negative role in the influence of a physician's profile photo information on affective trust.	PPI × PE→AT	–0.068	1.013	.16	No
H2c: Patient expertise will play a positive role in the impact of patient-generated information on cognitive trust.	PGI × PE→CT	0.127	1.730	.04	Yes
H2d: Patient expertise will play a positive role in the impact of patient-generated information on affective trust.	PGI × PE→AT	0.161	2.251	.01	Yes
H3a: The patient's cognitive trust in a physician will positively influence his or her choice of physician.	CT→CP[g]	0.437	7.553	<.001	Yes
H3b: The patient's affective trust in a physician will positively influence his or her choice of physician.	AT→CP	0.502	8.696	<.001	Yes

[a]PNI: physicians' nonprofile photo information.

[b]CT: cognitive trust.

[c]PPI: physicians' profile photo information.

[d]AT: affective trust.

[e]PGI: patient-generated information.

[f]PE: patient expertise.

[g]CP: choice of physician.

Figure 3. Moderating effect of patient expertise (PE) on cognitive trust (CT) by patient-generated information (PGI).

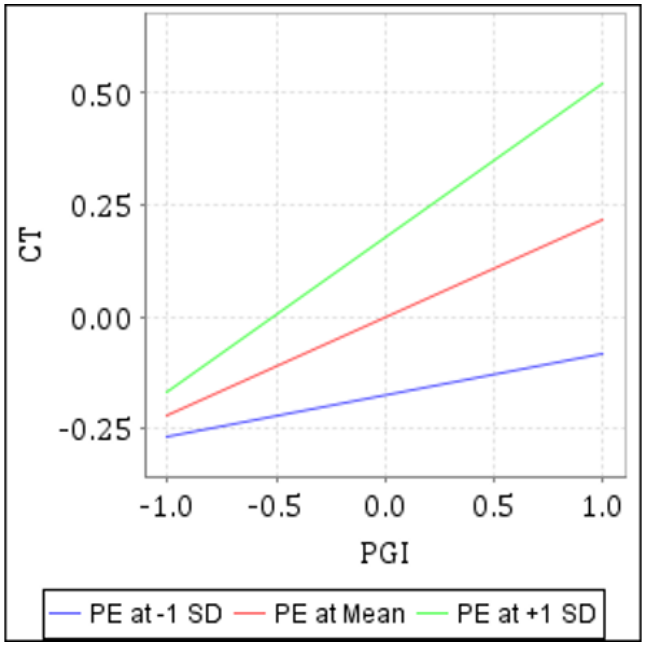
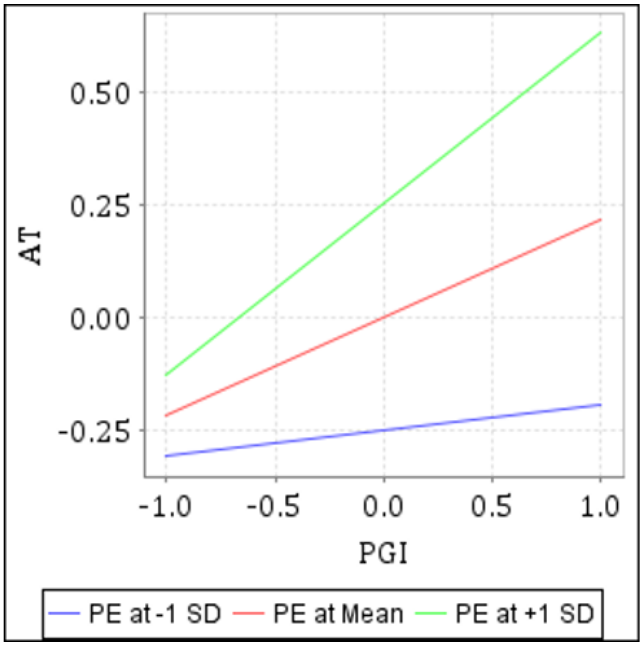


Figure 4. Moderating effect of patient expertise (PE) on affective trust (AT) by patient-generated information (PGI).



Discussion

Principal Findings

Based on the SOR conceptual framework, this paper explored the impact of different types of physician information on an mHealth app on patients' choice of physician. In the era of Network 2.0, a large amount of data will be delivered to consumers at little cost; with it comes cognitive load, so consumers are often willing to take shortcuts when making assessments or decisions [64]. Therefore, we believe that, although there is an increasing amount of physician information on mHealth apps, this is part of the information that plays a key role in patients' choice of physician. Such information—physicians' nonprofile photo information, physicians' profile photo information, and patient-generated information—influences patients' choice of physician through cognitive trust and affective trust and is regulated by patient expertise. In order to support our hypotheses, two studies were designed.

First, based on the capacity of human short-term memory, in Study 1 we discovered seven aspects of physician information that play a key role in patients' choice of physician through an eye-tracking experiment: physicians' profile photo, hospital, title, favorability rate, consultation numbers, satisfaction, and fees. Through this, we discovered an interesting phenomenon. Most of the participants subjectively believed that physicians' profile photos were not very useful when making their choices; however, our preliminary evidence showed that the influence of the visual information from the physicians' profile photos was beyond what they believed. In the field of online medical treatment, the influence of physicians' profile photos on patients' choice of physician has not been systematically discussed. Previous studies have suggested that smiling may increase consumers' perceived trust, and the attractiveness and perceived trustworthiness of profile photos may be related to each other, leading to the *beauty premium* phenomenon [54,65,66]. Based on this, we believe that physicians' profile photos are an influential factor that play a very important role in patients' choice of physician in mHealth services; this may create some potential economic benefits. Therefore, future research on visual information should be expanded.

Second, in Study 2 we found that physicians' profile photo information and physicians' nonprofile photo information positively influenced patients' choice of physician through affective trust and cognitive trust, respectively, which was not surprising. However, patient-generated information positively affected patients' choice of physician through cognitive trust and affective trust; patient expertise played a significant moderating role, that is, the higher the expertise of patients, the greater the role of patient-generated information. mHealth services include a process of interaction between physicians and patients. When patients make a choice, they not only pay attention to physicians' personal information, but are also influenced by other patients' suggestions to a large extent. Previous studies investigated the impact of user-generated information but generally concentrated on books, movies, and other consumer goods [67,68]; there is little literature on medical services. Intangible, heterogeneous service quality is often more difficult to evaluate than product quality [69]; this study fills the research gap in the medical field. Moreover, we believe that with the continuing development of the Internet, mHealth care will be further popularized and people's patient expertise regarding mHealth care will be improved; this means that patient-generated information will play a greater role.

We discovered another interesting phenomenon from our study results; affective trust played a more important role on patients' choice of physician than did cognitive trust. This supports the theory proposed by Komiak and Benbasat that in the field of e-commerce, affective trust plays a more important role than cognitive trust in determining consumers' willingness to adopt new services [70]. This discovery could be positive for some junior physicians. Nowadays, the competition for talent is

becoming more fierce; some physicians are often unable to secure employment in higher-quality hospitals or they fail to receive higher-ranked titles because of their relatively fewer qualifications. Although these may result in lower cognitive trust from patients, these physicians can make efforts in the direction of affective trust. This can include actively improving their personal photos by looking more formal and sporting professional smiles to increase their visual credibility. In addition, physicians can devote more energy to mHealth services, treat each patient with the utmost care and professionalism, and allow patients to benefit from their goodwill in order to improve patient-generated information, which could help improve their competitiveness in mHealth services.

Implications

The results of this study may also contribute to the development of mHealth apps; mHealth apps can be designed and improved upon in the following three aspects.

First, the patients' eye-tracking behavior on the mHealth platform and questionnaire results can be combined to determine the most efficient and effective way to sort information within apps. Then, information can be arranged according to the phenomenon that, in eye-tracking experiments, a user's attention has been shown to decrease from top to bottom and from left to right. According to this study, patients paid the most attention to favorability rate and the consultation numbers, so information such as this can be placed at the top left.

Second, mHealth platforms should actively improve the relevant mechanism of patient-generated information, considering its influence on patient trust; this would allow it to impact and promote the successful establishment of physician-patient relationships.

Third, mHealth platforms can authenticate patient expertise by setting up questionnaires. When setting up a physician recommendation list, platforms should consider giving more weight to patient-generated information (eg, favorability rate, consultation numbers, and satisfaction) when patient expertise is high. When patient expertise is low, platforms should consider giving more weight to physicians' profile photo information as well as their hospital and title information.

Limitations

The results from this study help in understanding the process of information selection, the usage of mHealth apps, and information that influences patients' choice regarding medical services. However, there are still some limitations that will need be addressed in future research. First, only 12 types of physician information were selected in this study; however, other information, such as patient comments and physician-patient conversation records, may also have an impact on patients' choice of physician. Future research can address the impact of these and other types of information. Second, mHealth care could play a relatively prominent role among the elderly [16], however, the participants of this study were younger (ie, 20-25 years of age in Study 1 and mainly under 30 years of age in Study 2). Follow-up research could include participants in other age groups in order to promote the widespread use of mHealth apps. In addition, the questionnaire survey method used in this study had certain subjective limitations. Follow-up research could apply the in-depth case study method, as it has advantages over the questionnaire method when analyzing dynamic phenomena [71].

Acknowledgments

This work was supported by the National Natural Science Foundation of China (#71971008 and #71371025), the Beijing Natural Science Foundation (#9182010), the Fundamental Research Funds for the Central Universities (#YWF-19-BJ-W-47), and the Teaching Reform Fund of Beihang University (#4302060).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Physician information levels.
[PDF File (Adobe PDF File), 116 KB - mhealth_v7i10e15544_app1.pdf]

Multimedia Appendix 2
Demographic characteristics of the participants.
[PDF File (Adobe PDF File), 120 KB - mhealth_v7i10e15544_app2.pdf]

References

1. Laugesen J, Hassanein K, Yuan Y. The impact of Internet health information on patient compliance: A research model and an empirical study. J Med Internet Res 2015 Jun 11;17(6):e143 [FREE Full text] [doi: 10.2196/jmir.4333] [Medline: 26068214]
2. Lester RT, van der Kop M, Taylor D, Alasaly K, Coleman J, Marra F. BC Medical Journal. 2011 Jun. M-health: Connecting patients to improve population and public health URL: https://www.bcmj.org/bccdc/m-health-connecting-patients-improve-population-and-public-health [accessed 3019-10-17]

3. Gunter R, Fernandes-Taylor S, Mahnke A, Awoyinka L, Schroeder C, Wiseman J, et al. Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113 [FREE Full text] [doi: 10.2196/mhealth.6023] [Medline: 27683059]
4. Tian M, Zhang J, Luo R, Chen S, Petrovic D, Redfern J, et al. mHealth interventions for health system strengthening in China: A systematic review. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e32 [FREE Full text] [doi: 10.2196/mhealth.6889] [Medline: 28302597]
5. Zhang H, Zhang H, Wang X, Yang Z, Zhao Y. Analysis of requirements for developing an mHealth-based health management platform. JMIR Mhealth Uhealth 2017 Aug 03;5(8):e117 [FREE Full text] [doi: 10.2196/mhealth.5890] [Medline: 28778840]
6. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. Electron Commer Res Appl 2016 Mar;16:55-65. [doi: 10.1016/j.elerap.2015.11.001]
7. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: Systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
8. Boulos MN, 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]
9. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: A decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
10. Willcox JC, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of women and health professionals on mHealth lifestyle interventions in pregnancy: A qualitative investigation. JMIR Mhealth Uhealth 2015 Oct 28;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]
11. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: Exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275. [doi: 10.1177/1460458214555041] [Medline: 25385165]
12. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: A participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
13. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): Development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
14. Guo X, Han X, Zhang X, Dang Y, Chen C. Investigating mHealth acceptance from a protection motivation theory perspective: Gender and age differences. Telemed J E Health 2015 Aug;21(8):661-669. [doi: 10.1089/tmj.2014.0166] [Medline: 25919800]
15. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: Empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
16. Meng F, Guo X, Peng Z, Lai K, Zhao X. Investigating the adoption of mobile health services by elderly users: Trust transfer model and survey study. JMIR Mhealth Uhealth 2019 Jan 08;7(1):e12269 [FREE Full text] [doi: 10.2196/12269] [Medline: 30622092]
17. Gao G, Greenwood BN, McCullough J, Agarwal R. A digital soapbox? The information value of online physician ratings. In: Proceedings of the Academy of Management Annual Meeting. 2012 Jul Presented at: Academy of Management Annual Meeting; August 3-7, 2012; Boston, MA p. 16354. [doi: 10.5465/AMBPP.2012.16354abstract]
18. Grabner-Kräuter S, Waiguny MK. Insights into the impact of online physician reviews on patients' decision making: Randomized experiment. J Med Internet Res 2015 Apr 09;17(4):e93 [FREE Full text] [doi: 10.2196/jmir.3991] [Medline: 25862516]
19. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
20. Swar B, Hameed T, Reychav I. Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search. Comput Human Behav 2017 May;70:416-425. [doi: 10.1016/j.chb.2016.12.068]
21. Lu N, Wu H. Exploring the impact of word-of-mouth about physicians' service quality on patient choice based on online health communities. BMC Med Inform Decis Mak 2016 Nov 26;16(1):151 [FREE Full text] [doi: 10.1186/s12911-016-0386-0] [Medline: 27888834]
22. Corritore CL, Wiedenbeck S, Kracher B, Marble RP. Online trust and health information websites. Int J Technol Hum Interact 2012 Mar;8(4):92-115. [doi: 10.4018/jthi.2012100106]
23. Ziemba P, Watróbski J, Karczmarczyk A, Jankowski J, Wolski W. Integrated approach to e-commerce websites evaluation with the use of surveys and eye tracking-based experiments. In: Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS). 2017 Sep Presented at: Federated Conference on Computer Science and Information Systems (FedCSIS); September 3-6, 2017; Prague, Czech Republic p. 1019-1030. [doi: 10.15439/2017F320]
24. Hoffmann CP, Lutz C, Meckel M. Digital natives or digital immigrants? The impact of user characteristics on online trust. J Manag Inf Syst 2015 Mar 09;31(3):138-171. [doi: 10.1080/07421222.2014.995538]
25. Ghoreishi S, Mohammadi S. The World of Computer Science and Information Technology Journal. 2015 Aug 11. Analysis of factors influence online trust: A unified model URL: http://wcsit.org/volume-5-issue-9-2015 [accessed 2018-08-11]

26. Swan JE, Nolan JJ. Gaining customer trust: A conceptual guide for the salesperson. J Pers Selling Sales Manage 1985 Nov;5(2):39-48. [doi: 10.1080/08853134.1985.10754400]
27. Singh J, Sirdeshmukh D. Agency and trust mechanisms in consumer satisfaction and loyalty judgments. J Acad Mark Sci 2000 Jan 01;28(1):150-167. [doi: 10.1177/0092070300281014]
28. McAllister DJ. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manage J 1995 Feb 01;38(1):24-59. [doi: 10.2307/256727]
29. Lewis JD, Weigert A. Trust as a social reality. Soc Forces 1985 Jun 01;63(4):967-985. [doi: 10.1093/sf/63.4.967]
30. Blomqvist K. The many faces of trust. Scand J Manage 1997 Sep;13(3):271-286. [doi: 10.1016/s0956-5221(97)84644-1]
31. Pavlou P, Dimoka A. Understanding and mitigating product uncertainty in online auction marketplaces. In: Proceedings of the Industry Studies Conference. 2008 Mar Presented at: Industry Studies Conference; May 1-2, 2008; Boston,MA. [doi: 10.2139/ssrn.1135006]
32. Pearson SD, Raeke LH. Patients' trust in physicians: Many theories, few measures, and little data. J Gen Intern Med 2000 Jul;15(7):509-513 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.11002.x] [Medline: 10940139]
33. Calefato F, Lanubile F, Novielli N. The role of social media in affective trust building in customer-supplier relationships. Electron Commer Res 2015 Jul 16;15(4):453-482. [doi: 10.1007/s10660-015-9194-3]
34. Mehrabian A, Russell JA. An Approach to Environmental Psychology. Cambridge, MA: The MIT Press; 1974.
35. Bitner MJ. Servicescapes: The impact of physical surroundings on customers and employees. J Mark 2018 Nov 28;56(2):57-71. [doi: 10.1177/002224299205600205]
36. Jacoby J. Stimulus-organism-response reconsidered: An evolutionary step in modeling (consumer) behavior. J Consum Psychol 2002 Dec;12(1):51-57. [doi: 10.2307/1480568]
37. Eroglu SA, Machleit KA, Davis LM. Empirical testing of a model of online store atmospherics and shopper responses. Psychol Mark 2003 Jan 17;20(2):139-150. [doi: 10.1002/mar.10064]
38. Eckel CC, Petrie R. Face value. Am Econ Rev 2011 Jun;101(4):1497-1513. [doi: 10.1257/aer.101.4.1497]
39. Rule NO, Ambady N, Adams RB, Ozono H, Nakashima S, Yoshikawa S, et al. Polling the face: Prediction and consensus across cultures. J Pers Soc Psychol 2010 Jan;98(1):1-15. [doi: 10.1037/a0017673] [Medline: 20053027]
40. Pavlou PA, Gefen D. Building effective online marketplaces with institution-based trust. Inf Syst Res 2004 Mar;15(1):37-59. [doi: 10.1287/isre.1040.0015]
41. Dellarocas C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Manage Sci 2003 Oct;49(10):1407-1424. [doi: 10.1287/mnsc.49.10.1407.17308]
42. Eysenbach G. Credibility of health information and digital media: New perspectives and implications for youth. In: Metzger MJ, Flanagin AJ, editors. Digital Media, Youth, and Credibility. Cambridge, MA: The MIT Press; Jan 2008:123-154.
43. Yang H, Zhang X. Investigating the effect of paid and free feedback about physicians' telemedicine services on patients' and physicians' behaviors: Panel data analysis. J Med Internet Res 2019 Mar 22;21(3):e12156 [FREE Full text] [doi: 10.2196/12156] [Medline: 30900997]
44. Gao G, Greenwood BN, Agarwal R, McCullough JS. Vocal minority and silent majority: How do online ratings reflect population perceptions of quality? MIS Q 2015 Sep;39(3):565-590. [doi: 10.2139/ssrn.2629837]
45. Park DH, Kim S. The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. Electron Commer Res Appl 2008 Dec;7(4):399-410. [doi: 10.1016/j.elerap.2007.12.001]
46. Zou P, Yu B, Hao Y. Does the valence of online consumer reviews matter for consumer decision making? The moderating role of consumer expertise. J Comput 2011 Mar 17;6(3):484-488 [FREE Full text] [doi: 10.4304/jcp.6.3.484-488]
47. Cheung C, Xiao B, Liu I. The impact of observational learning and electronic word of mouth on consumer purchase decisions: The moderating role of consumer expertise and consumer involvement. In: Proceedings of the 45th Hawaii International Conference on System Sciences. 2012 Jan Presented at: 45th Hawaii International Conference on System Sciences; January 4-7, 2012; Maui, HI. [doi: 10.1109/hicss.2012.570]
48. Gefen D, Benbasat I, Pavlou P. A research agenda for trust in online environments. J Manag Inf Syst 2014 Dec 08;24(4):275-286. [doi: 10.2753/MIS0742-1222240411]
49. Boutla M, Supalla T, Newport EL, Bavelier D. Short-term memory span: Insights from sign language. Nat Neurosci 2004 Sep 15;7(9):997-1002 [FREE Full text] [doi: 10.1038/nn1298] [Medline: 15311279]
50. Liu X, Guo X, Wu H, Vogel D. Doctor's effort influence on online reputation and popularity. In: Proceedings of the International Conference on Smart Health (ICSH). Cham, Switzerland: Springer International Publishing; 2014 Jul Presented at: International Conference on Smart Health (ICSH); July 10-11, 2014; Beijing, China p. 233-244. [doi: 10.1007/978-3-319-08416-9_11]
51. Yang H, Guo X, Wu T, Ju X. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. Electron Commer Res Appl 2015 May;14(3):192-203. [doi: 10.1016/j.elerap.2015.04.001]
52. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The effect of online effort and reputation of physicians on patients' choice: 3-wave data analysis of China's Good Doctor website. J Med Internet Res 2019 Mar 08;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]

53. Henderson JM, Williams CC, Falk RJ. Eye movements are functional during face learning. Mem Cognit 2005 Jan;33(1):98-106. [doi: 10.3758/BF03195300]
54. Ert E, Fleischer A, Magen N. Trust and reputation in the sharing economy: The role of personal photos in Airbnb. Tour Manag 2016 Aug;55:62-73. [doi: 10.1016/j.tourman.2016.01.013]
55. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: An integrative typology. Inf Syst Res 2002 Sep;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]
56. Liu C, Marchewka JT, Lu J, Yu C. Beyond concern: A privacy-trust-behavioral intention model of electronic commerce. Inf Manage 2005 Jan;42(2):289-304. [doi: 10.1016/j.im.2004.01.003]
57. Just MA, Carpenter PA. Eye fixations and cognitive processes. Cogn Psychol 1976 Oct;8(4):441-480. [doi: 10.1016/0010-0285(76)90015-3]
58. Chin WW. The partial least squares approach to structural equation modeling. In: Marcoulides GA, editor. Modern Methods For Business Research. New York, NY: Psychology Press; 2013:295-336.
59. Yi MY, Davis FD. Developing and validating an observational learning model of computer software training and skill acquisition. Inf Syst Res 2003 Jun;14(2):146-169. [doi: 10.1287/isre.14.2.146.16016]
60. SmartPLS. URL: http://www.smartpls.com [accessed 2015-08-16]
61. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
62. Hair JF, Sarstedt M, Ringle CM, Mena JA. An assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Mark Sci 2011 Jun 7;40(3):414-433. [doi: 10.1007/s11747-011-0261-6]
63. Hair JF, Hult GTM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd edition. Thousand Oaks, CA: SAGE Publications; 2017.
64. Fiske ST, Taylor SE. Social Cognition. New York, NY: McGraw-Hill; 1991.
65. Scharlemann JP, Eckel CC, Kacelnik A, Wilson RK. The value of a smile: Game theory with a human face. J Econ Psychol 2001 Oct;22(5):617-640. [doi: 10.1016/S0167-4870(01)00059-9]
66. Wilson RK, Eckel CC. Judging a book by its cover: Beauty and expectations in the trust game. Polit Res Q 2016 Jul 02;59(2):189-202. [doi: 10.1177/106591290605900202]
67. Dellarocas C, Gao G, Narayan R. Are consumers more likely to contribute online reviews for hit or niche products? J Manag Inf Syst 2014 Dec 08;27(2):127-158. [doi: 10.2753/mis0742-1222270204]
68. Archak N, Ghose A, Ipeirotis PG. Deriving the pricing power of product features by mining consumer reviews. Manage Sci 2011 Aug;57(8):1485-1509. [doi: 10.1287/mnsc.1110.1370]
69. Parasuraman A, Zeithaml VA, Berry LL. A conceptual model of service quality and its implications for future research. J Mark 2018 Nov 19;49(4):41-50. [doi: 10.1177/002224298504900403]
70. Komiak SYX, Benbasat I. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q 2006 Dec;30(4):941-960. [doi: 10.2307/25148760]
71. Shi Y, Sia CL, Chen H. Leveraging social grouping for trust building in foreign electronic commerce firms: An exploratory study. Int J Inf Manage 2013 Jun;33(3):419-428. [doi: 10.1016/j.ijinfomgt.2013.02.001]

Abbreviations

AOI: area of interest
AT: affective trust
AVE: average variance extraction
CB-SEM: Covariance-Based Structural Equation Modeling
CP: choice of physician
CT: cognitive trust
H: hypothesis
mHealth: mobile health
PE: patient expertise
PGI: patient-generated information
PLS-SEM: Partial Least Squares Structural Equation Modeling
PNI: physicians' nonprofile photo information
PPI: physicians' profile photo information
SOR: stimulus-organism-response

Edited by G Eysenbach; submitted 20.07.19; peer-reviewed by S Hugh-Jones, M Tkac; comments to author 13.08.19; revised version received 02.09.19; accepted 30.09.19; published 23.10.19.

Please cite as:
Shan W, Wang Y, Luan J, Tang P
The Influence of Physician Information on Patients' Choice of Physician in mHealth Services Using China's Chunyu Doctor App: Eye-Tracking and Questionnaire Study
JMIR Mhealth Uhealth 2019;7(10):e15544
URL: http://mhealth.jmir.org/2019/10/e15544/
doi:10.2196/15544
PMID:31647466

Original Paper

Pulse Rate Variability in Emergency Physicians During Shifts: Pilot Cross-Sectional Study

Gregory Andrew Peters[1,2*], BA; Matthew L Wong[1,2*], MPH, MD; Joshua W Joseph[1,2], MS, MD; Leon D Sanchez[1,2], MPH, MD

[1]Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[2]Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Gregory Andrew Peters, BA
Department of Emergency Medicine
Beth Israel Deaconess Medical Center
1 Deaconess Rd
Boston, MA, 02215
United States
Phone: 1 6177542339
Email: greg.peters826@gmail.com

Abstract

Background: The high prevalence of physician burnout, particularly in emergency medicine, has garnered national attention in recent years. Objective means of measuring stress while at work can facilitate research into stress reduction interventions, and wearable photoplethysmography (PPG) technology has been proposed as a potential solution. However, the use of low-burden wearable biosensors to study training and clinical practice among emergency physicians (EP) remains untested.

Objective: This pilot study aimed to (1) determine the feasibility of recording on-shift photoplethysmographic data from EP, (2) assess the quality of these data, and (3) calculate standard pulse rate variability (PRV) metrics from the acquired dataset and examine patterns in these variables over the course of an academic year.

Methods: A total of 21 EP wore PPG biosensors on their wrists during clinical work in the emergency department during a 9-hour shift. Recordings were collected during the first quarter of the academic year, then again during the fourth quarter of the same year for comparison. The overall rate of usable data collection per time was computed. Standard pulse rate (PR) and PRV metrics from these two time points were calculated and entered into Student *t* tests.

Results: More than 400 hours of data were entered into these analyses. Interpretable data were captured during 8.54% of the total recording time overall. In the fourth quarter of the academic year compared with the first quarter, there was no significant difference in median PR (75.8 vs 76.8; *P*=.57), mean R-R interval (0.81 vs 0.80; *P*=.32), SD of R-R interval (0.11 vs 0.11; *P*=.93), root mean square of successive difference of R-R interval (0.81 vs 0.80; *P*=.96), low-frequency power (3.5×103 vs 3.4×103; *P*=.79), high-frequency power (8.5×103 vs 8.3×103; *P*=.91), or low-frequency to high-frequency ratio (0.42 vs 0.41; *P*=.43), respectively. Power estimates for each of these tests exceeded .90. A secondary analysis of the resident-only subgroup similarly showed no significant differences over time, despite power estimates greater than .80.

Conclusions: Although the use of PPG biosensors to record real-time physiological data from EP while providing clinical care seems operationally feasible, this study fails to support the notion that such an approach can efficiently provide reliable estimates of metrics of interest. No significant differences in PR or PRV metrics were found at the end of the year compared with the beginning. Although these methods may offer useful applications to other domains, it may currently have limited utility in the contexts of physician training and wellness.

(JMIR Mhealth Uhealth 2019;7(10):e13909) doi:10.2196/13909

KEYWORDS

emergency medicine; burnout; photoplethysmography; emergency physicians; physician wellness; stress; heart rate variability; pulse rate variability

Introduction

The concept of burnout in the workplace, roughly defined as mental exhaustion because of chronic work-related stress, was first introduced nearly 50 years ago [1]. Once applied to physicians, the field of physician burnout research has steadily grown at a rapid rate [2], especially as it pertains to medical training in the United States [3]. A recent study of more than 16,000 US residents showed that the majority of their sample reported burnout [4]. In particular, emergency medicine (EM) has been identified by this body of research as a relatively high-risk specialty, as emergency physicians (EPs) frequently show the highest rates of burnout [5,6]. As physician burnout continues to increase in prevalence and attracts greater attention both within and outside medicine, efforts to alleviate burnout among medical providers have become a priority [7-10]. A natural corollary from this movement has been an internationally concerted search for new, objective ways to measure burnout, to serve as a complementary approach to the current body of research predominated by surveys and self-report inventories such as the Maslach Burnout Inventory [11].

In this search for new technologies to direct efforts to relieve physician burnout, heart rate (HR) and heart rate variability (HRV) metrics have recently shown promise in multiple applications to medical providers [12-17]. Clinical applications of HRV analysis trace roots back more than a century ago [18], and the combination of improved technology and increased interest in physician well-being have spurred this new area of investigation [19]. Despite concerns regarding the validity of certain applications of HRV data (eg, the use of low-frequency [LF] power of HRV—and in turn ratio of LF to high-frequency [HF] power—to measure sympathovagal balance has been criticized [20-23]), a large body of evidence supports the notion that HRV analysis can illuminate the balance between sympathetic and parasympathetic tone in the body [24], enabling the use of HRV to study autonomic responses to mental stress [25]. A particularly noteworthy study of HRV in surgeons highlights the value of objective measures of stress among medical providers, reporting that acute care surgeons in their sample showed levels of physiologic stress that were elevated out of proportion to self-reported stress [17]. These results offer compelling evidence that physicians, especially those routinely exposed to traumatic situations in highly time-sensitive settings, might provide unreliable assessments of their own stress. Thus, widespread use of objective, ecologically valid metrics of stress might represent a key piece of the puzzle in efforts to understand and alleviate burnout among EPs.

While HRV recorded by electrocardiography (ECG) offers a seemingly effective means by which to chart physiologic manifestations of stress responses in small samples of providers for limited durations, widespread application of this methodology to routine monitoring of physicians and trainees at work remains unfeasible. However, measurement of pulse rate (PR) and pulse rate variability (PRV) via wearable photoplethysmography (PPG) biosensor technology has been posed as an extremely low-burden alterative to ECG measurement with significantly greater potential for scalability. In a basic sense, while ECG measures changes in electrical activity in the heart, PPG measures changes in blood volume pulse (BVP) in the peripheral vasculature. The context of this physiologic relationship lies in the assumption that chronic exposures to workplace stressors that contribute to burnout are psychologically experienced by EPs, corresponding to neurological registration of the experience that triggers an autonomic response. This autonomic response, expressed as a change in sympathovagal balance, directly influences cardiac function as measurable via HR and HRV and that these changes in cardiac function will alter systemic blood flow and be reflected in PR and PRV metrics. Therefore, in an effort to measure stress responses reflected in autonomic changes that influence cardiac activity, ECG measures these phenomena more proximally than PPG, while PPG captures more downstream physiologic variables that are vulnerable to more physiologic interference and require reliance on a greater number of assumptions (eg, effects of the respiratory cycle on blood flow and multifactorial variance in systemic vascular resistance, [26] as a well-known, extreme example of uncoupling between ECG and PPG within this conceptual chain, consider the pathological state of pulseless electrical activity). The adequacy of PRV as a surrogate for HRV remains controversial [26-31], and to the knowledge of this author's group, no studies have been published to evaluate the suitability of PRV analysis to study physician wellness.

This study aimed to evaluate the feasibility of collecting and analyzing PPG data from EPs while they work and to investigate whether patterns in their PR and PRV changed over the course of an academic year. To achieve this aim, PPG data collected from resident and attending EM physicians during the fourth quarter of the year were compared with an analogous PPG dataset from the same providers during the first quarter of the same year. This approach is founded on the notion that as trainees gain experience in their role over the course of a year, perception of stressors and their physiologic manifestations among the trainees—as well as among the attending physicians directly responsible for them—should decrease. For example, it has been shown that the odds that an individual EP will register a detectable concentration of salivary cortisol after completing a shift significantly decrease over the course of a year [32]. An additional analysis will be restricted to the resident-only subgroup to investigate whether patterns in PR and PRV may be specific to EM trainees. Given the novelty of the methods employed in this pilot study, a secondary aim involved assessing whether currently available PPG technology is ready for application to the ecologically valid study of physician wellness (ie, real-time examination of autonomic manifestations of stress at work), embedded within the context of a comprehensive literature review. In summary, the essential purpose of this pilot study is to answer the question: Can wearable PPG sensors offer a scalable, objective, and ecologically valid method to study workplace stress among EPs?

Methods

Institutional Review

This study was approved by the Institution's Committee on Clinical Investigations and the Department of Emergency Medicine's Medical Education Executive Committee.

Study Setting and Population

Participants were recruited from an academic ED in late June 2016 and 2017 at the beginning of the academic year to support 2 iterations of the study in consecutive years. Following an introductory information session during a scheduled department conference, 3 emails were sent to all EM residents and attending physicians inviting them to participate in the study. Volunteers who responded to these recruitment emails underwent an informed consent procedure, individually conducted by a trained research assistant. No financial compensation was offered for participation. All recordings took place at a single site: the ED at the host academic medical center.

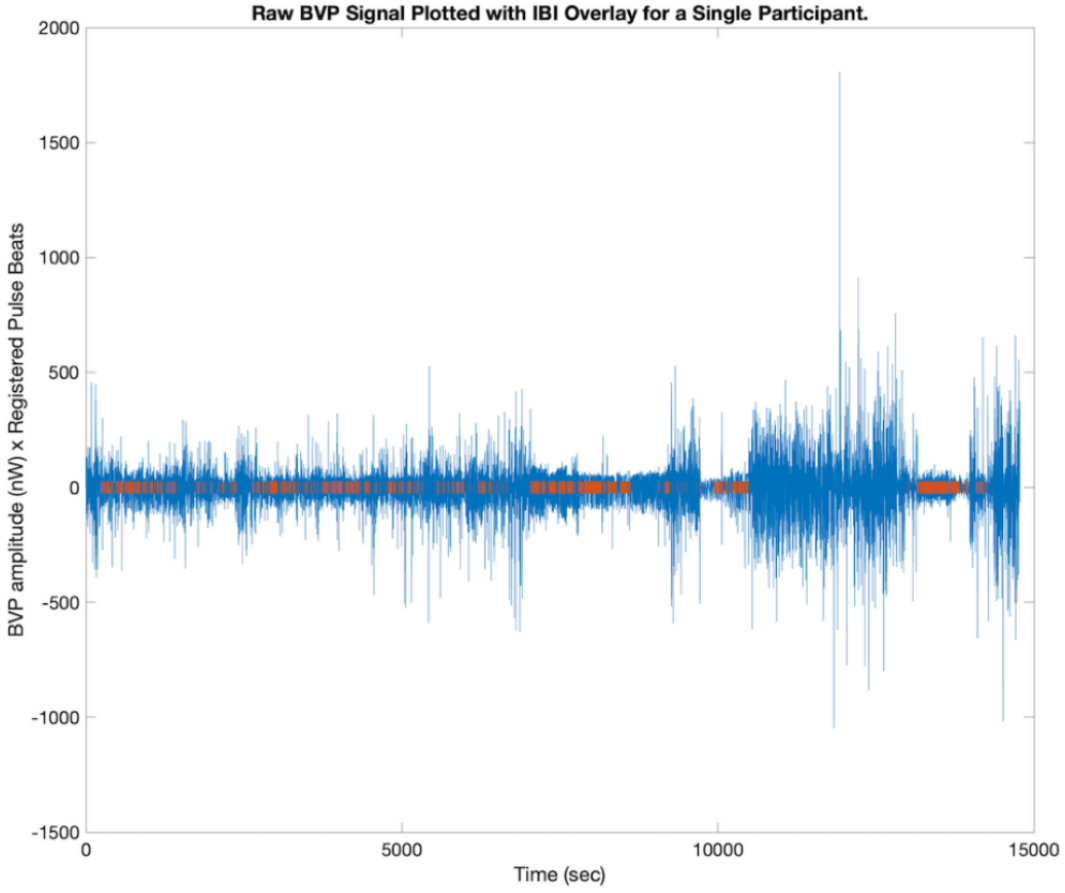
Data Collection

Participants wore an Empatica E4 PPG biosensor watch (Empatica Inc) during at least 1 shift in the ED at the host institution during the first quarter (Q1: July-September) and final quarter (Q4: April-June) of 1 academic year. Continuous PPG data were recorded over the course of the entire 9-hour shift (14:00-23:00) while performing routine clinical care. Each recording automatically generated 2 forms of data used for this study: a raw PPG signal (BVP data) and an interbeat interval (IBI) log, each measured with respect to time at a sampling frequency of 64 Hz. Compilation of these datasets relied upon standard detection and filtering algorithms developed by Empatica [33,34].

Data Filtering and Inclusion Criteria

Once compiled, each recording entered a screening process before being included in the study. Data filtering and analysis procedures were executed via a self-developed set of MATLAB scripts designed to accomplish the following operations (Mathworks). First, the timing of each PPG recording was compared with the timing of each shift to verify that each included dataset reflects the participant's physiology while providing care in the ED. Second, the total number of heart beats registered in each IBI dataset was screened for a minimum total of 300 beats. Given that each IBI dataset is generated via the application of a standard heart beat detection algorithm that excludes segments of recording consistent with artifacts or low signal to noise ratio, this screening measure served to ensure sufficient clean signal density in each recording. Third, each recording that met these inclusion criteria was plotted in a PPG versus time graph, and additional markers were superimposed on the x-axis to indicate time periods included in the IBI logs (Figure 1).

Figure 1. Raw blood volume pulse (BVP) amplitude measured in nanowatts from a single PPG recording plotted in blue, with registered pulse beats from the corresponding interbeat interval (IBI) log marked in red. PPG: photoplethysmography.



These plots were visually inspected to add a final opportunity to manually detect segments of included IBI data sourced from unreliable PPG signal. Finally, to sustain a repeated-measures analysis, data were only included from participants with 1 satisfactory recording in both Q1 and Q4. Once included, each PPG and IBI dataset was entered into 2 primary analyses, one in the time domain and the other in the frequency domain, providing complementary assessments of PR and PRV over time on a repeated-measures basis.

Data Analysis

Each IBI dataset was entered into a time-domain analysis designed according to previously validated standards [35]. Variables of interest included median PR, mean R-R interval, standard deviation of R-R interval, and root mean square of successive difference of R-R interval (RMSSD). Calculations of RMSSD were performed by isolating the single longest stretch of consecutive heart beats detected within each IBI log. The PPG dataset yielded calculations of power in the low-frequency range (LF: 0.05-0.15 Hz) and high-frequency range (HF: 0.15-0.40 Hz), as well as LF:HF ratio (LHR), composing the frequency-domain analysis [35].

Recordings dated from Q1 and Q4 were separated into 2 time points for each participant. The previously described standard measures of PR and PRV were calculated for each recording, and Q1 versus. Q4 was compared via paired Student *t* tests for each metric. The data used in this study satisfy the requisite assumptions upon which paired-samples *t* tests rely. Statistical power for these tests was calculated using a publicly available MATLAB algorithm [36].

Toward the aim to assess the quality of the data collected through these methods, the IBI log from each of the original recordings included in the analysis was transformed into a time vector that only included 1-second segments of time during which pulse beats were detected (including the interval between consecutive beats), according to the standard Empatica algorithm. This vector was then transposed upon a second vector representing the total recording time captured in each corresponding PPG dataset, sampled at the same frequency (graphically represented in Figure 1). Finally, a quotient of time during which pulse beats were detected over the grand total recording time, in seconds, was computed to determine the overall percent of data that was ultimately interpretable with respect to time.

Results

A total of 51 EPs contributed 95 recordings, producing over 800 hours of data. Following application of the previously described screening procedure, roughly 440 hours of data from 21 EPs were included in the study (Table 1). Interpretable PPG data constituted 8.54% of the total recording time undertaken in this study. A total of 10 participants were trainees and 11 were attending physicians. Two participants were included in both years of the study. Five of the 21 participants were female, including 2 of the 10 residents. Residents from all years of training were represented, including 4 postgraduate year (PGY)-1, 2 PGY-2, and 4 PGY-3 trainees. The mean age of participants at the time of enrollment was 32.8 years, with 29.9 years and 35.6 years for residents and attending physicians, respectively.

Table 1. Descriptive characteristics of participants included in the study.

Characteristic	Value
Sample size (N=21)	
Resident physicians, n (%)	10 (48)
Attending physicians, n (%)	11 (52)
Sex (female), n (%)	
Overall sample (N=21)	5 (24)
Resident physicians (n=10)	2 (20)
Attending physicians (n=11)	3 (27)
Age (years), mean (SD)	
Overall sanmple (N=21)	32.8 (5.9)
Resident physicians (n=10)	29.9 (3.9)
Attending physicians (n=11)	35.9 (6.3)
Trainees	
Total	10 (100)
PGY[a]-1, n (%)	4 (40)
PGY-2, n (%)	2 (20)
PGY-3, n (%)	4 (40)

[a]PGY: postgraduate year.

Time-domain analysis revealed no significant changes in PR or PRV in Q4 compared with Q1. Table 2 details the descriptive data from the primary analysis. Median PR in Q4 was 75.8 (SD 13.5) compared with 76.8 (SD 10.8) in Q1 (*P*=.57). Comparisons of PRV measures between Q4 and Q1 also did not reach statistical significance, including group mean RMSSD of 0.81 (SD 0.16) versus 0.80 (SD 0.11), respectively (*P*=.96). Similarly, frequency-domain analysis did not show any differences in Q4 versus Q1, including a comparison of LHR in which group means were 0.42 (SD 0.04) and 0.41 (SD 0.03), respectively (*P*=.43). These negative findings were observed despite adequate power, with estimated power (1−ß) values exceeding .90 for each of the 8 Student *t* tests included in this study, corresponding to the 8 PR and PRV measures of interest used, as outlined in Table 3.

Table 2. Mean (SD) values of measures in primary analysis including all participants (XeY denotes scientific notation).

Quarter	MPR[a]	MRRI[b]	SDNN[c]	RMSSD[d]	LF[e]	HF[f]	LHR[g]
Quarter 1	76.8 (10.8)	0.80 (0.113)	0.11 (0.035)	0.80 (0.113)	3.4e+3 (1.7e+3)	8.3e+3 (4.2e+3)	0.41 (0.04)
Quarter 4	75.8 (13.5)	0.81 (0.137)	0.11 (0.023)	0.81 (0.161)	3.5e+3 (2.1e+3)	8.5e+3 (5.3e+3)	0.42 (0.03)

[a]MPR: median pulse rate (pulse beats per minute).

[b]MRRI: mean R-R interval (seconds per pulse beat).

[c]SDNN: standard deviation of R-R interval.

[d]RMSSD: root mean square of successive difference of R-R interval.

[e]LF: low-frequency power (ms^2).

[f]HF: high-frequency power (ms^2).

[g]LHR: ratio of LF to HF.

Table 3. Results of primary analysis including all participants (Student t tests).

Metric	MPR[a]	MRRI[b]	SDNN[c]	RMSSD[d]	LF[e]	HF[f]	LHR[g]
P value	.57	.32	.93	.96	.79	.91	.43
Power (1−ß)	.94	.92	.95	.95	.94	.95	.90

[a]MPR: median pulse rate (pulse beats per minute).

[b]MRRI: mean R-R interval (seconds per pulse beat).

[c]SDNN: standard deviation of R-R interval.

[d]RMSSD: root mean square of successive difference of R-R interval.

[e]LF: low-frequency power (ms^2).

[f]HF: high-frequency power (ms^2).

[g]LHR: ratio of LF to HF.

A secondary analysis of the resident-only subgroup similarly failed to reveal any significant changes in PR or PRV in Q4 compared with Q1, despite estimated power values greater than .80. Tables 4 and 5 detail results from this secondary analysis.

Table 4. Mean (SD) values of measures in primary analysis including trainees only (XeY denotes scientific notation).

Quarter	MPR[a]	MRRI[b]	SDNN[c]	RMSSD[d]	LF[e]	HF[f]	LHR[g]
Quarter 1	78.5 (10.5)	0.78 (0.095)	0.12 (0.04)	0.79 (0.123)	3.6e+3 (1.7e+3)	8.8e+3 (4.2e+3)	0.41 (0.02)
Quarter 4	74.7 (12.5)	0.82 (0.123)	0.11 (0.022)	0.82 (0.119)	3.6e+3 (2.5e+3)	8.7e+3 (6.2e+3)	0.42 (0.02)

[a]MPR: median pulse rate (pulse beats per minute).

[b]MRRI: mean R-R interval (seconds per pulse beat).

[c]SDNN: standard deviation of R-R interval.

[d]RMSSD: root mean square of successive difference of R-R interval.

[e]LF: low-frequency power (ms^2).

[f]HF: high-frequency power (ms^2).

[g]LHR: ratio of LF to HF.

Table 5. Results of secondary analysis including trainees only (Student t tests).

Metric	MPR[a]	MRRI[b]	SDNN[c]	RMSSD[d]	LF[e]	HF[f]	LHR[g]
P value	.28	.15	.44	.49	.99	.95	.16
Power (1−ß)	.88	.83	.93	.92	.95	.95	.81

[a]MPR: median pulse rate (pulse beats per minute).

[b]MRRI: mean R-R interval (seconds per pulse beat).

[c]SDNN: standard deviation of R-R interval.

[d]RMSSD: root mean square of successive difference of R-R interval.

[e]LF: low-frequency power (ms^2).

[f]HF: high-frequency power (ms^2).

[g]LHR: ratio of LF to HF.

A post hoc analysis of the original 95 recordings demonstrated a yield of 8.54% with respect to time, meaning less than one-tenth of the PPG data recorded contained useful substrate for the PR and PRV analyses employed in this study.

Discussion

Principal Findings

This study demonstrates the feasibility of recording real-time physiological data in EPs while they practice in an operational sense but fails to support the notion that the current state of wearable PPG biosensor technology can efficiently and reliably measure variables of interest in a meaningful way. The evidence produced by this study consistently supports the null hypothesis that there is no difference in PR or PRV patterns among EPs at the end of an academic year compared with the beginning, or more precisely, fails to support the alternative hypothesis that such a difference exists. The primary analysis failed to detect any differences over time in EPs at large, and the secondary analysis demonstrated a similar inability to find any significant changes among EM residents. These negative findings were observed in both arms of the study despite the use of complementary time-domain and frequency-domain analyses and despite adequate power calculated for each analysis. Given the anticipated effect of the passing academic year to decrease physiologic stress—including attending physicians working with more experienced and autonomous trainees, and particularly for residents with the benefit of 6 to 10 months of training in their assigned role—the question of whether this methodology can capture such effects requires careful consideration. Therefore, the aim of this pilot study to assess the suitability of these methods for these purposes emerges as paramount.

The closest precedent for this study is a 1998 study that used ECG recordings in 12 EPs to report increased sympathetic tone during night shifts [37]. Clearly, additional research is needed to develop objective, ecologically valid metrics of wellness among EPs, ideally using tools that can be upscaled to make widespread participation possible even while providing care in the ED. As previously mentioned, no literature exists on the use of PPG to study physician wellness—in any specialty—to the knowledge of this author's group. However, a wealth of research continues to focus on evaluating the adequacy of PPG technology to approximate the HRV analyses that have demonstrated promise in this field [13-15]. This body of research comparing the 2 methodologies has yielded mixed results. The literature seems to weigh in favor of the accuracy of PPG measurement for healthy human subjects at rest [27]. On the other hand, evidence consistently calls into question the ability of PPG-enabled wearable biosensors to achieve useful recordings in human subjects undergoing active tasks marked by interference from factors such as motion artefact, inconsistent skin contact, and increased degrees of respiratory-cardiovascular interaction among other sources of physiologic variability [26,28,38]. Of course, these conditions apply to the study of EPs providing care in the ED and likely account for the yield of less than 10% usable data from our total recording time. In addition, further ambiguity surrounds more specific applications of HRV concepts to PRV methodology, such as translation of frequency domain metrics from ECG to PPG data [29,30], and even more fundamentally, use of the frequency domain to assess sympathetic activity in either ECG or PPG [20-23].

Taking this body of the literature into consideration, this study adds several new pieces of information. The primary goal of this study was to investigate whether any patterns in PR and PRV among EPs could be detected by comparing PPG data collected during Q4 versus Q1 of an academic year. These results indicate that no such patterns were found, regardless of the domain of analysis or training status of participants included and despite adequate power to conduct each test. The secondary question that emerges from this conclusion is whether the novel methodology of this study possesses the ability to detect changes in physiology linked to stress if they do in fact exist.

On the basis of a focused review of the pertinent literature, it seems unclear at best whether wearable PPG sensor technology has yet been developed to the extent necessary to characterize stress levels among EPs, especially while providing care in the ED. In particular, the quality assessment analysis demonstrating a yield of useful data less than 10% suggests that this methodology is at least an inefficient, if not inaccurate, means of probing physiologic changes in EPs at work. Taken together, the results of this study and existing body of the literature seem to suggest that PPG technology has not yet matured to the degree required for the use of PRV analysis in this field of research. However, it remains crucial to understand that wearable PPG sensors lie at the center of an intensely active area of multidisciplinary research and technological development [39].

For example, a new singular spectrum analysis–based method to automate detection and correction of artefacts in ECG data was published in 2019 following the completion of this study; such advancements might significantly enhance the quality of data that is readily accessible to researchers without a background in the technical aspects of this field [40]. Although this study suggests that current PPG-enabled wearable biosensors technology might have limited utility in the study of physician wellness in real time, advancements in the collection, analysis, and interpretation of PPG data will likely transform the field in the near future and warrant further investigation.

Limitations

Consideration of future developments surrounding this methodology aside, further research in the short term is needed to study sources of burnout in EPs. For example, several limitations of this study will need to be addressed in future investigations. First, as previously discussed, the current state of wearable PPG sensor technology remains highly vulnerable to interference from settings such as the ED. Perhaps, studies of smaller scope using HRV must be used to investigate EP wellness until PRV analysis has reached requisite maturity for these purposes. Alternatively, perhaps a chest strap can be used as an intermediate wearable biosensor in future studies to reduce interference inherent to PPG measurement at the wrist without subjecting participants to mobile ECG recording. Second, although the repeated-measures design of this study was chosen to enhance sensitivity to detect changes within participants over time, it unfortunately limited sample size, especially among trainees. Resident EM physicians at the host academic hospital frequently rotate at outside ED locations, significantly limiting opportunities to capture recordings within both time windows. Perhaps, an alternative design with greater sample size might uncover significant patterns in PRV among EPs, although the power calculations performed in this study call such considerations into question. Third, in an effort to test whether this methodology can offer a time-sensitive, ecologically valid way to study EPs at work, the exact duration of recordings entered into these analyses was not precisely standardized, which can pose issues regarding the frequency-domain analyses in particular. Instead, participants were simply instructed to record PPG throughout their entire 9-hour shift, and each recording that met inclusion criteria as detailed in the Methods section was entered into the frequency-domain analysis in its entirety. Similarly, the devices used in this study are capable of recording accelerometer data that can be used to inform exclusion of PPG data associated with significant motion. This measure was not taken for the purpose of this study because of the concern that exclusion of times at which EPs are active might lead to the systematic loss of the stressors this pilot study aimed to capture, but future studies of this kind must carefully weigh the benefits and disadvantages of this decision with respect to the primary goals of the study.

Fourth, this study lacked a direct comparison between HRV and PRV. The primary goal of this study was to evaluate a novel tool to measure stress among EPs in real time to guide efforts aimed at alleviating physician burnout. As a result, the focus of this study was to assess the utility of PRV analysis, given its readiness for widespread application to resident and attending EM physicians, rather than to validate the technology used for these purposes against another that bears limited ability to serve this primary goal. However, further research is recommended to use ECG alongside PPG to compare HRV versus PRV analyses in EPs providing care in the ED. Such studies can more directly assess the still unclear question as to whether the physiologic changes investigated in this study exist but cannot be detected by PRV analysis versus the possibility that these changes simply do not exist. Similarly, this study did not use self-reported data as a comparator, although evidence suggests that caution must be used when relying upon the assumption that different modalities of multifactorial concepts such as stress should necessarily correlate with one another [17]. Future studies will be required to address these limitations and further advance the search for new objective tools that can characterize burnout among EPs.

Conclusions

The results of this study do not indicate any changes in PR or PRV among EPs over the course of an academic year. Due to the limitations of this study, it remains unclear whether such patterns exist and merely went undetected, although calculations suggest this study was sufficiently powered. The finding that less than one-tenth of the recording time dedicated to this study yielded useful substrate for PR and PRV analyses lends support to the notion that PPG technology might not yet be ready for application to these purposes, suggesting that the methods used in this study are poorly suited to test the hypothesis in question. A review of the literature supports the suggestion from this study that wearable PPG sensor technology has not yet matured to the extent required for accurate measurement of physiologic reflection of mental stress among EPs while at work on a large scale. However, active research and development surrounding this technology will likely offer new opportunities for investigation in the near future. Further research is required to identify new tools that can inform mounting efforts to alleviate physician burnout.

In summary, this study adds the first evaluation of wearable PPG biosensor technology as an ecologically valid, objective measure of workplace-related stress among EPs to the literature. Although this methodology proved to be low-burden and therefore easily scalable, data quality was a prohibitive issue. At the least, the current state of wearable PPG biosensors is subject to technological limitations that render it unable to reliably measure PR and PRV in active EPs. At the most, PPG methodology is subject to physiologic interference that precludes study of upstream concepts such as sympathovagal balance as a reflection of stress among physically active participants, such as EPs during shifts. In conclusion, this pilot study suggests that alternative methods must be explored to establish an objective, scalable, ecologically valid way to measure stress among EPs at work.

Conflicts of Interest

None declared.

References

1. Freudenberger HJ. Staff burn - out. J Soc Issues 2010;30(1):159-165. [doi: 10.1111/j.1540-4560.1974.tb00706.x]
2. Eckleberry-Hunt J, Kirkpatrick H, Barbera T. The problems with burnout research. Acad Med 2018 Mar;93(3):367-370. [doi: 10.1097/ACM.0000000000001890] [Medline: 28817432]
3. Ishak WW, Lederer S, Mandili C, Nikravesh R, Seligman L, Vasa M, et al. Burnout during residency training: a literature review. J Grad Med Educ 2009 Dec;1(2):236-242 [FREE Full text] [doi: 10.4300/JGME-D-09-00054.1] [Medline: 21975985]
4. West CP, Shanafelt TD, Kolars JC. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. J Am Med Assoc 2011 Sep 7;306(9):952-960. [doi: 10.1001/jama.2011.1247] [Medline: 21900135]
5. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 8;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
6. Shanafelt TD, Hasan O, Dyrbye LN, Sinsky C, Satele D, Sloan J, et al. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. Mayo Clin Proc 2015 Dec;90(12):1600-1613. [doi: 10.1016/j.mayocp.2015.08.023] [Medline: 26653297]
7. Shanafelt TD, Dyrbye LN, West CP. Addressing physician burnout: the way forward. J Am Med Assoc 2017 Mar 7;317(9):901-902. [doi: 10.1001/jama.2017.0076] [Medline: 28196201]
8. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. Mayo Clin Proc 2017 Jan;92(1):129-146. [doi: 10.1016/j.mayocp.2016.10.004] [Medline: 27871627]
9. National Academy of Medicine. 2018. Action Collaborative on Clinician Well-Being and Resilience URL: https://nam.edu/initiatives/clinician-resilience-and-well-being/ [accessed 2019-02-22] [WebCite Cache ID 76NpH1a6N]
10. American College of Physicians. 2019. Physician Well-Being and Professional Satisfaction URL: https://www.acponline.org/practice-resources/physician-well-being-and-professional-satisfaction [accessed 2019-02-22] [WebCite Cache ID 76NplA9zz]
11. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory Manual. Norwood, MA: Consulting Psychologists; 1981.
12. Langelotz C, Scharfenberg M, Haase O, Schwenk W. Stress and heart rate variability in surgeons during a 24-hour shift. Arch Surg 2008 Aug;143(8):751-755. [doi: 10.1001/archsurg.143.8.751] [Medline: 18711034]
13. Malmberg B, Persson R, Flisberg P, Ørbaek P. Heart rate variability changes in physicians working on night call. Int Arch Occup Environ Health 2011 Mar;84(3):293-301. [doi: 10.1007/s00420-010-0593-4] [Medline: 21107598]
14. Hernández-Gaytan SI, Rothenberg SJ, Landsbergis P, Becerril LC, de León-León G, Collins SM, et al. Job strain and heart rate variability in resident physicians within a general hospital. Am J Ind Med 2013 Jan;56(1):38-48. [doi: 10.1002/ajim.22098] [Medline: 22886873]
15. Amirian I, Andersen LT, Rosenberg J, Gögenur I. Decreased heart rate variability in surgeons during night shifts. Can J Surg 2014 Oct;57(5):300-304. [doi: 10.1503/cjs.028813] [Medline: 25265102]
16. Rieger A, Stoll R, Kreuzfeld S, Behrens K, Weippert M. Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress. Int Arch Occup Environ Health 2014 Feb;87(2):165-174. [doi: 10.1007/s00420-013-0847-z] [Medline: 23370764]
17. Joseph B, Parvaneh S, Swartz T, Haider AA, Hassan A, Kulvatunyou N, et al. Stress among surgical attending physicians and trainees: a quantitative assessment during trauma activation and emergency surgeries. J Trauma Acute Care Surg 2016 Oct;81(4):723-728. [doi: 10.1097/TA.0000000000001162] [Medline: 27389128]
18. Eppinger H, Hess L, Kraus WM, Jelliffe SE. Vagotonia: A Clinical Study in Vegetative Neurology. Vagotonia: Nervous and Mental Disease Publishing Company; 1917.
19. Billman GE. Heart rate variability - a historical perspective. Front Physiol 2011;2:86 [FREE Full text] [doi: 10.3389/fphys.2011.00086] [Medline: 22144961]
20. Billman GE. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. Front Physiol 2013;4:26 [FREE Full text] [doi: 10.3389/fphys.2013.00026] [Medline: 23431279]
21. Goldstein DS, Bentho O, Park MY, Sharabi Y. Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes. Exp Physiol 2011 Dec;96(12):1255-1261 [FREE Full text] [doi: 10.1113/expphysiol.2010.056259] [Medline: 21890520]
22. del Paso GA, Langewitz W, Mulder LJ, van Roon A, Duschek S. The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies. Psychophysiology 2013 May;50(5):477-487. [doi: 10.1111/psyp.12027] [Medline: 23445494]
23. Goedhart AD, Willemsen G, Houtveen JH, Boomsma DI, de Geus EJ. Comparing low frequency heart rate variability and preejection period: two sides of a different coin. Psychophysiology 2008 Nov;45(6):1086-1090. [doi: 10.1111/j.1469-8986.2008.00710.x] [Medline: 18823417]

24. Malliani A, Pagani M, Lombardi F, Cerutti S. Cardiovascular neural regulation explored in the frequency domain. Circulation 1991 Aug;84(2):482-492. [doi: 10.1161/01.cir.84.2.482] [Medline: 1860193]
25. Castaldo R, Melillo P, Bracale U, Caserta M, Triassi M, Pecchia L. Acute mental stress assessment via short term HRV analysis in healthy adults: a systematic review with meta-analysis. Biomed Signal Process Control 2015 Apr;18:370-377. [doi: 10.1016/j.bspc.2015.02.012]
26. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 2013 Jun 5;166(1):15-29. [doi: 10.1016/j.ijcard.2012.03.119] [Medline: 22809539]
27. Pinheiro N, Couceiro R, Henriques J, Muehlsteff J, Quintal I, Goncalves L, et al. Can PPG be used for HRV analysis? Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:2945-2949. [doi: 10.1109/EMBC.2016.7591347] [Medline: 28268930]
28. Pietilä J, Mehrang S, Tolonen J, Helander E, Jimison H, Pavel M, et al. Evaluation of the Accuracy and Reliability for Photoplethysmography Based Heart Rate and Beat-to-Beat Detection During Daily Activities. In: Proceedings of the European Medical and Biological Engineering Confernce. 2017 Presented at: EMBEC'17; June 11-15, 2017; Tampere, Finland p. 145-148. [doi: 10.1007/978-981-10-5122-7_37]
29. Iozzia L, Cerina L, Mainardi L. Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA. Physiol Meas 2016 Nov;37(11):1934-1944. [doi: 10.1088/0967-3334/37/11/1934] [Medline: 27681456]
30. Dehkordi P, Garde A, Karlen W, Wensley D, Ansermino JM, Dumont GA. Pulse rate variability compared with heart rate variability in children with and without sleep disordered breathing. Conf Proc IEEE Eng Med Biol Soc 2013;2013:6563-6566. [doi: 10.1109/EMBC.2013.6611059] [Medline: 24111246]
31. Wong JS, Lu WA, Wu KT, Liu M, Chen GY, Kuo CD. A comparative study of pulse rate variability and heart rate variability in healthy subjects. J Clin Monit Comput 2012 Apr;26(2):107-114. [doi: 10.1007/s10877-012-9340-6] [Medline: 22350367]
32. Wong ML, Peters G, Joseph JW, Chung A, Sanchez LD. Salivary cortisol concentrations, grit, and the effect of time. AEM Educ Train 2019 May 20;26(2):107-114. [doi: 10.1002/aet2.10353]
33. Empatica Support. 2018. How is IBI.csv Obtained? URL: http://support.empatica.com/hc/en-us/articles/201912319-How-is-IBI-csv-obtained- [accessed 2019-06-11]
34. Empatica Support. 2019. What Should I Know to Use the PPG/IBI Data in My Experiment? URL: http://support.empatica.com/hc/en-us/articles/203621335-What-should-I-know-to-use-the-PPG-IBI-data-in-my-experiment- [accessed 2019-06-11]
35. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
36. MathWorks. 2019. Sample Size and Power of Test - MATLAB sampsizepwr URL: https://www.mathworks.com/help/stats/sampsizepwr.html [accessed 2019-06-10]
37. Adams SL, Roxe DM, Weiss J, Zhang F, Rosenthal JE. Ambulatory blood pressure and Holter monitoring of emergency physicians before, during, and after a night shift. Acad Emerg Med 1998 Sep;5(9):871-877 [FREE Full text] [doi: 10.1111/j.1553-2712.1998.tb02816.x] [Medline: 9754499]
38. Giardino ND, Lehrer PM, Edelberg R. Comparison of finger plethysmograph to ECG in the measurement of heart rate variability. Psychophysiology 2002 Mar;39(2):246-253. [doi: 10.1017/S0048577202990049] [Medline: 12212675]
39. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable photoplethysmographic sensors—past and present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]
40. Lang M. Automatic near real-time outlier detection and correction in cardiac interbeat interval series for heart rate variability analysis: singular spectrum analysis-based approach. JMIR Biomed Eng 2019 Jan 30;4(1):e10740. [doi: 10.2196/10740]

Abbreviations

BVP: blood volume pulse
ECG: electrocardiography
EM: emergency medicine
EP: emergency physicians
HF: high frequency
HR: heart rate
HRV: heart rate variability
IBI: interbeat interval
LF: low frequency
LHR: ratio of LF to HF
MPR: median pulse rate (pulse beats per minute)
MRRI: mean R-R interval (seconds per pulse beat)
PGY: postgraduate year
PPG: photoplethysmography

PR: pulse rate
PRV: pulse rate variability
RMSSD: root mean square of successive difference of R-R interval
SDNN: standard deviation of R-R interval

Edited by G Eysenbach; submitted 04.03.19; peer-reviewed by M Lang, IN Gomez, C Molbech; comments to author 28.04.19; revised version received 20.06.19; accepted 07.07.19; published 02.10.19.

Please cite as:
Peters GA, Wong ML, Joseph JW, Sanchez LD
Pulse Rate Variability in Emergency Physicians During Shifts: Pilot Cross-Sectional Study
JMIR Mhealth Uhealth 2019;7(10):e13909
URL: https://mhealth.jmir.org/2019/10/e13909
doi:10.2196/13909
PMID:31579017

Original Paper

Validation and Acceptability of a Cuffless Wrist-Worn Wearable Blood Pressure Monitoring Device Among Users and Health Care Professionals: Mixed Methods Study

Sheikh Mohammed Shariful Islam[1*], MBBS, MPH, PhD, FESC; Susie Cartledge[1], PhD; Chandan Karmakar[2], PhD; Jonathan Charles Rawstorn[1], PhD; Steve F Fraser[1], PhD; Clara Chow[3*], MBBS, PhD; Ralph Maddison[1*], PhD

[1]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Faculty of Health, Deakin University, Burwood, Australia

[2]School of IT, Faculty of Science, Engineering and Built Environment, Deakin University, Burwood, Australia

[3]Charles Perkins Centre Westmead, University of Sydney, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Sheikh Mohammed Shariful Islam, MBBS, MPH, PhD, FESC
Institute for Physical Activity and Nutrition
School of Exercise and Nutrition Sciences, Faculty of Health
Deakin University
221 Burwood Highway
Burwood, VIC 3125
Australia
Phone: 61 451733373
Email: shariful@deakin.edu.au

Abstract

Background: Blood pressure (BP) is an important modifiable cardiovascular risk factor, yet its long-term monitoring remains problematic. Wearable cuffless devices enable the capture of multiple BP measures during everyday activities and could improve BP monitoring, but little is known about their validity or acceptability.

Objective: This study aimed to validate a wrist-worn cuffless wearable BP device (Model T2; TMART Technologies Limited) and assess its acceptability among users and health care professionals.

Methods: A mixed methods study was conducted to examine the validity and comparability of a wearable cuffless BP device against ambulatory and home devices. BP was measured simultaneously over 24 hours using wearable and ambulatory devices and over 7 days using wearable and home devices. Pearson correlation coefficients compared the degree of association between the measures, and limits of agreement (LOA; Bland-Altman plots) were generated to assess measurement bias. Semistructured interviews were conducted with users and 10 health care professionals to assess acceptability, facilitators, and barriers to using the wearable device. Interviews were audio recorded, transcribed, and analyzed.

Results: A total of 9090 BP measurements were collected from 20 healthy volunteers (mean 20.3 years, SD 5.4; N=10 females). Mean (SD) systolic BP (SBP)/diastolic BP (DBP) measured using the ambulatory (24 hours), home (7 days), and wearable (7 days) devices were 126 (SD 10)/75 (SD 6) mm Hg, 112 (SD 10)/71 (SD 9) mm Hg and 125 (SD 4)/77 (SD 3) mm Hg, respectively. Mean (LOA) biases and precision between the wearable and ambulatory devices over 24 hours were 0.5 (−10.1 to 11.1) mm Hg for SBP and 2.24 (−17.6 to 13.1) mm Hg for DBP. The mean biases (LOA) and precision between the wearable and home device over 7 days were −12.7 (−28.7 to 3.4) mm Hg for SBP and −5.6 (−20.5 to 9.2) mm Hg for DBP. The wearable BP device was well accepted by participants who found the device easy to wear and use. Both participants and health care providers agreed that the wearable cuffless devices were easy to use and that they could be used to improve BP monitoring.

Conclusions: Wearable BP measures compared well against a gold-standard ambulatory device, indicating potential for this user-friendly method to augment BP management, particularly by enabling long-term monitoring that could improve treatment titration and increase understanding of users' BP response during daily activity and stressors.

(JMIR Mhealth Uhealth 2019;7(10):e14706) doi:10.2196/14706

KEYWORDS

hypertension; cardiovascular disease; wearable device; blood pressure; ambulatory blood pressure monitoring

Introduction

High blood pressure (BP) or hypertension is the leading risk factor for cardiovascular disease, including myocardial infarction, stroke, and kidney disease, and accounts for 14% (10.4 million) of total deaths globally [1,2]. In 2015, an estimated 874 million people had high BP, but the disease burden is estimated to be much greater as high BP remains undetected, untreated, and uncontrolled in many individuals [2,3]. Over the last three decades, the incidence of high BP has increased worldwide and is projected to increase further, mostly because of an aging population, urbanization, reduced physical activity, and unhealthy diet [2]. High BP also leads to significant declines in productivity and economic burden and remains as a significant public health challenge.

Accurate BP measurement is essential for the diagnosis and management of hypertension; however, current measurement methods are suboptimal [4]. Measurements taken by health care professionals during medical consultations can be inaccurate because of *white coat* hypertension, and infrequent measurements may not represent typical hemodynamics [5]. Home BP measurements have been recommended by several guidelines [5-7], but it is impractical to measure BP during daily bouts of activity and at night, which has prognostic and therapeutic importance [8]. Ambulatory BP measurement devices can provide frequent valid measures across a 24-hour period and are the gold standard for clinical use, but they are not appropriate for long-term monitoring because they are intrusive, cumbersome, and costly [9].

In recent years, a number of cuffless wearable devices have emerged that enable frequent and unobtrusive BP measurement throughout the user's usual everyday activities. Wearable cuffless BP devices are broadly defined as those worn on/attached to the body and without a pneumatic cuff, and they overcome many limitations of traditional ambulatory devices. Therefore, these devices may be suitable for collecting regular BP measurements over prolonged periods. Regular long-term monitoring could enable more comprehensive assessment of BP status and treatment adherence [10] as long as measurement accuracy meets guideline recommendations [11,12]. In a comprehensive literature review, we identified a number of commercially available cuffless, continuous, wearable BP monitoring devices [10] and selected a wrist-worn device that could be used in day-to-day life. However, before recommending the use of this device, it must be validated against standard BP measurement devices. At the same time, as these devices have only recently become available as consumer-grade products for BP measurement, little is known about the users' and health care professionals' perspective about their use in real life. Therefore, the primary aim of this study was to validate a wrist-worn cuffless wearable BP device against a gold-standard ambulatory BP device. Secondary aims of this study were to compare the wearable BP measurements against a home BP device and assess the acceptability of a wearable device among end users and health care professionals.

Methods

Study Design

A mixed methods (quantitative and qualitative) approach was used in this study.

Participants

We recruited a convenience sample of 20 healthy volunteers via advertisements and flyers at the Deakin University Burwood campus and selected general practice clinics in Melbourne, Australia. We included adults (aged ≥18 years) with normal BP (<140/90 mm Hg) who were willing to wear an ambulatory device for 24 hours, a wearable device for 7 days, and record home BP 3 times per day for 7 days. Participants with high BP (>140/90 mm Hg), serious medical conditions, and limited mobility and those taking BP medication at baseline were excluded. In addition, we purposively recruited 10 health care professionals who manage patients with high BP and had clinical experience with a range of BP devices—including cardiologists, general practitioners, nurses, pharmacists, and exercise physiologists—to ascertain the barriers and facilitators of wearable devices and acceptability for use in their clinical practice.

Ethics

Written informed consent was obtained from all participants at the time of enrollment. The study was approved by the Deakin University Faculty of Health Human Ethics Advisory Group (HEAG-H 135_2017).

Data Collection and Variables

A research assistant was trained in data collection procedures, device configuration, testing, and operation for 1 week at the Deakin University. Data were collected from October 2017 to April 2018. Data regarding socioeconomic status (age, sex, education, employment, occupation, and income), self-reported comorbidities, smoking, alcohol use, cognitive function, physical activity (light, moderate, and vigorous activities/times per week for ≥15 min), diet (fruits and vegetable consumption: servings/week), and medication were collected face-to-face using a standardized questionnaire. Weight and height were measured at enrollment. Body mass index was calculated as weight in kilograms divided by height in meters squared.

Blood Pressure Measurements

Baseline BP was measured using an Omron automated device (Omron HEM 7121; Omron Corp) at the time of enrollment. A total of 3 baseline measures were obtained. The first measurement was discarded, and the mean of the remaining 2 readings was calculated following standard practice [13]. Starting at the time of enrollment, study participants were fitted with an ambulatory BP device (Model TM-2430; A & D Medical Corp) with appropriate cuff size, which is highly accurate and has been validated according to international standards and recommended for clinical use [14]. The ambulatory device was programmed to measure BP every 30

min during the day and every 60 min during the night for 24 hours. There were no restrictions of daily activities. Concurrently, study participants were asked to use the wearable cuffless device (Model T2; TMART Technologies Limited; Figure 1) in the nondominant hand, for example, a right-handed person was instructed to use the wearable device in the left hand. The wearable device measured BP every 60 min for 7 days. Participants were also provided with a home BP device (Omron HEM 7121) and a printed log sheet to measure and record BP 3 times per day—morning, afternoon, and evening, at consistent times self-selected by participants—for 7 days. All participants received instruction and demonstration about how to use each device at enrollment.

The wearable cuffless device was chosen as it is commercially available in Australia, is affordable (approximately Aus $50), is lightweight and easy to wear, and is waterproof and thus can be worn at all times. Participants were asked to wear the device on the wrist of their nondominant hand. This device uses MPU6500 sensors that do not require calibration before use, and it uses Bluetooth functionality to wirelessly send data to a mobile phone app (Wearfit) for storage and self-monitoring visualization. The device (minimum OS compatibility is Android 4.4 or iOS 8.0) also measured heart rate, blood oxygen saturation, sleep time, steps, mileage, and calories consumed among other features. However, these parameters were not considered in this study.

Data from the wearable and ambulatory devices were downloaded from the Wearfit app and onboard memory, respectively, after the participation period. Home BP measurements were recorded manually on a preformatted study log.

Figure 1. Wearable blood pressure device (T2 Mart blood pressure device).

Semistructured Interviews

Interviews were conducted with users at day 7 to determine the acceptability and usability of the wearable BP device. The interviews comprised questions about the overall experiences of using the provided BP measurement devices. Brief semistructured interviews were also conducted with health care professionals to determine acceptability as well as barriers and facilitators to real-world use. Health care professionals were briefed about the wearable BP device and how it worked. They were asked about the use of patient-reported BP data in their clinics and the future usability of wearable BP data. All interviews were audio recorded and transcribed verbatim.

Outcomes

The primary outcome was the mean difference between wearable and ambulatory BP measurements over 24 hours. Secondary outcomes included the mean difference between wearable and home BP measurements over 7 days, barriers and facilitators of using wearable BP devices by users, and the acceptability of wearable devices among a diverse group of health care professionals.

Data Analysis

Data were presented as mean, SD, and range. Wearable BP measurements were compared against corresponding reference ambulatory and home measurements (eg, Wearable BP device 1st measurement W1 vs Reference device 1st measurement R1; see Table 1). We manually investigated extreme values (5th, 10th, and 15th percentiles from both sides of the distribution) for all devices to check for possible measurement noise. As the data did not differ significantly, we used the full dataset from all devices in these analyses. Data were explored graphically using box plots and scatter plots. We considered subjective daily average between wearable and ambulatory devices. We also estimated daytime BP using measures recorded between 7 am and 9 pm. A $P<.05$ was considered statistically significant. Missing data were not imputed. Data analyses were performed using MATLAB 2017a software.

Table 1. Procedure for reference and wearable device blood pressure measurements validation.

BP[a] measurement	Number
Initial BP measurements[b]	
Take reference BP measurement (office BP)	R[c]0
Take wearable device BP measurement	W[d]0
Validation BP measurements for accuracy evaluation	
Take first reference BP measurement (mean 24-hour ambulatory BP monitoring)	R1
Take first wearable device BP measurement (mean 24-hour wearable device)	W1
Take second reference BP measurement (mean 7-day home BP monitoring)	R2
Take second wearable device BP measurement (mean 7-day wearable device)	W2

[a]BP: blood pressure.

[b]Measurement R0 was not used in the evaluation of reference BP distribution and variability criteria. Measurements R0 and W0 were not used in the evaluation of the test device accuracy.

[c]R: reference blood pressure device.

[d]W: wearable blood pressure device.

We used nonparametric Mann-Whitney U tests to compare the mean (SD) of the devices. Systolic and diastolic measurement biases were calculated as reference wearable measurement. We also assessed measurement accuracy by calculating the mean absolute difference (MAD) and mean absolute percentage differences (MAPD) between the devices [14]. The MAD and MAPD were calculated as follows: MAD= $(\Sigma^{n}_{i=1 |}\ p_i - y_{i |})/n$ and MAPD=$(\Sigma^{n}_{i=1 |}\ 100\ (p_i - y_{i |}\ y_i)/n$, where p_i and y_i are the average wearable and reference device measurements, respectively, and n is the sample size. Measurement accuracy was graded according to the following accepted clinical standards: grade A, MAD less than or equal to 5 mm Hg; grade B, MAD 5 to 6 mm Hg; grade C, MAD 6 to 7 mm Hg; and grade D, MAD greater than or equal to 7 mm Hg [15]. Relative reliability was estimated by calculating Pearson correlation coefficients to compare the degree of association [16]. Standardized Bland-Altman scatterplots and limits of agreement (LOA) were used to assess absolute reliability and the variability of measurement biases across the measurement range.

A content analysis was applied to the semistructured interview data [17]. This method is appropriate, given our aims to describe (1) usability and acceptability of the wearable device among participants and (2) acceptability of this method among health care professionals caring for people with hypertension. Once transcripts were read in their entirety, coding of main themes was performed manually, and a log was maintained in a spreadsheet form. Categories of code were developed from the data through an iterative process. Coding was initially performed separately for study participants and health care professionals to assess if there were consistent categories and themes between groups.

Results

Study Participants and Measurements

A total of 9090 systolic BP (SBP) and diastolic BP (DBP) data (1530 ambulatory, 6720 wearables, and 840 home devices) were analyzed. Participants' age was 20.3 (SD 5.4) years, half were female, and mean baseline BP was 112/74 mm Hg. Additional participants' characteristics are reported in Table 2.

Table 2. Characteristics of the study participants.

Characteristics	Study participants (n=20)
Male, n (%)	10 (50)
Age (years)	
Mean (SD)	20.3 (5.4)
Range	37.2-18.5
Body mass index	
Mean (SD)	23.6 (3.3)
Range	19.9-25.9
Married/living with partner, n (%)	8 (40)
Education (master's or above), n (%)	13 (65)
Employment (fulltime), n (%)	12 (60)
Baseline systolic BP[a]	
Mean (SD)	112.35 (9.79)
Range	95-131
Baseline diastolic BP	
Mean (SD)	73.75 (9.14)
Range	47-96

[a]BP: blood pressure.

Measurement Biases

Table 3 summarizes BP measured across devices; BP was similar for both ambulatory and wearable devices (SBP 126 [SD 10] vs 125 [SD 5]; DBP 75 [SD 6] vs 77 [SD 9]), and there were no statistically significant differences over 24 hours (P>.05). The MAD between wearable and ambulatory devices over 24 hours was less than 7 mm Hg for both SBP and DBP. The average 24-hour BP data obtained using the wearable and ambulatory device showed poor relationship (SBP: r=0.16, P=.51; DBP: r=−0.15, P=.53; Multimedia Appendix 1). Mean SBP and DBP measured by ambulatory and wearable BP devices did not differ significantly (P>.05; Figure 2). Figure 3 shows the Bland-Altman plots comparing wearable and ambulatory devices; the mean biases (LOA) was 0.5 (−10.1 to 11.1) mm Hg for SBP and 2.24 (−17.6 to 13.1) mm Hg for DBP. The mean difference (2 SDs) were 0.08 (2 SD 20.69) in SBP and 2.46 (2 SD 15.03) in DBP.

BP values differed significantly between wearable and home devices over 7 days (P≤.01; see Table 3). Mean (LOA) daytime biases for wearable versus home devices were −13.9 (−33.8 to 5.9) mm Hg for SBP and −6.4 (−24.6 to 11.8) mm Hg for DBP. Daytime 7-day mean (SD) SBP and DBP were 126 (SD 6) mm Hg and 78 (SD 4) mm Hg, respectively. Moreover, 7-day MAD was greater than 7 mm Hg for SBP and less than 7 mm Hg for DBP. The mean biases (LOA) between wearable and home devices over 7 days were −12.7 (−28.7 to 3.4) mm Hg for SBP and −5.6 (−20.5 to 9.2) mm Hg for DBP. Daytime biases were similar (SBP −13.9 [−33.8 to 5.9] mm Hg; DBP −6.4 [−24.6 to 11.8] mm Hg; Multimedia Appendices 2 and 3). Similar differences were observed in the MAPD between wearable and home devices.

Table 3. Blood pressure values measured by different devices.

BP[a]	Mean (SD)	Range	MD[b] (SD)	MAD[c] (SD)	MAPD[d] (SD)	Limits of agreement, MD (2SDs)	*P* value
Device worn for 24 hours							
Systolic BP			**0.08 (10.56)**	**7.63 (7.09)**	**6.01 (5.42)**	**0.08 (20.69)**	**>.05[e]**
Wearable	125 (5)	119-138					
Ambulatory	126 (10)	111-150					
Diastolic BP			**2.46 (7.67)**	**5.90 (5.34)**	**7.48 (6.45)**	**2.46 (15.03)**	**>.05[e]**
Wearable	77 (9)	72-87					
Ambulatory	75 (6)	64-90					
7 days							
Systolic BP			**13.19 (8.31)**	**13.19 (8.31)**	**10.56 (6.64)**	—	**<.01[f]**
Wearable	125 (4)	113-139					
Home	112 (10)	85-135					
Diastolic BP			**5.86 (7.62)**	**7.28 (6.21)**	**9.37 (8.04)**	—	**<.01[f]**
Wearable	77 (3)	68-87					
Home	71 (8)	50-90					

[a]BP: blood pressure.

[b]MD: mean difference.

[c]MAD: mean absolute difference.

[d]MAPD: mean absolute percentage difference.

[e]*P* for mean difference between ambulatory and wearable (24 hours) devices.

[f]*P* for mean difference between home (7 days) and wearable (7 days) devices.

Figure 2. Box plot of ambulatory and wearable systolic and diastolic BP measurements. *P* value calculated using nonparametric Mann-Whitney U Test. *P*>.05 indicates the absence of systematic measurement bias between devices. ABPM: ambulatory blood pressure monitoring; BP: blood pressure; WBPM: wearable blood pressure monitoring device.

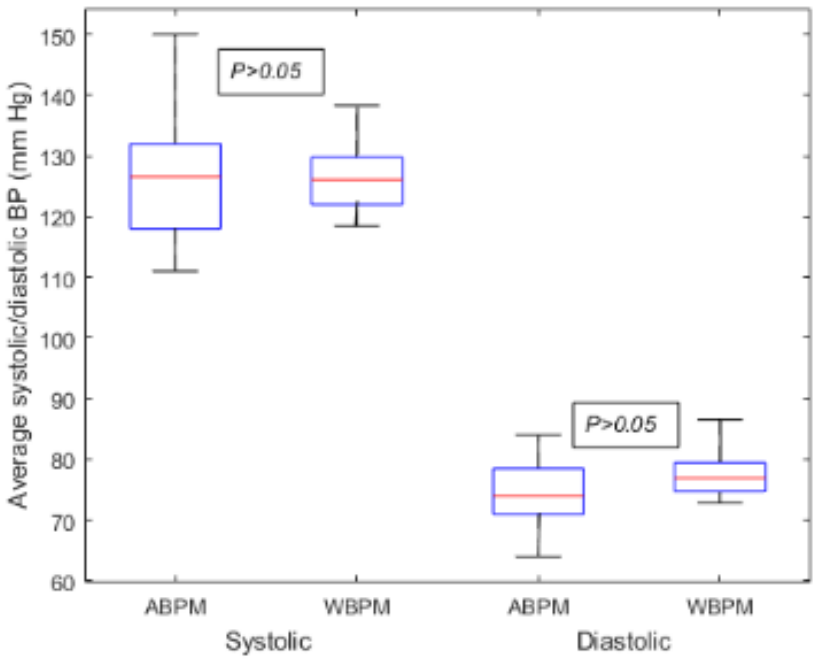
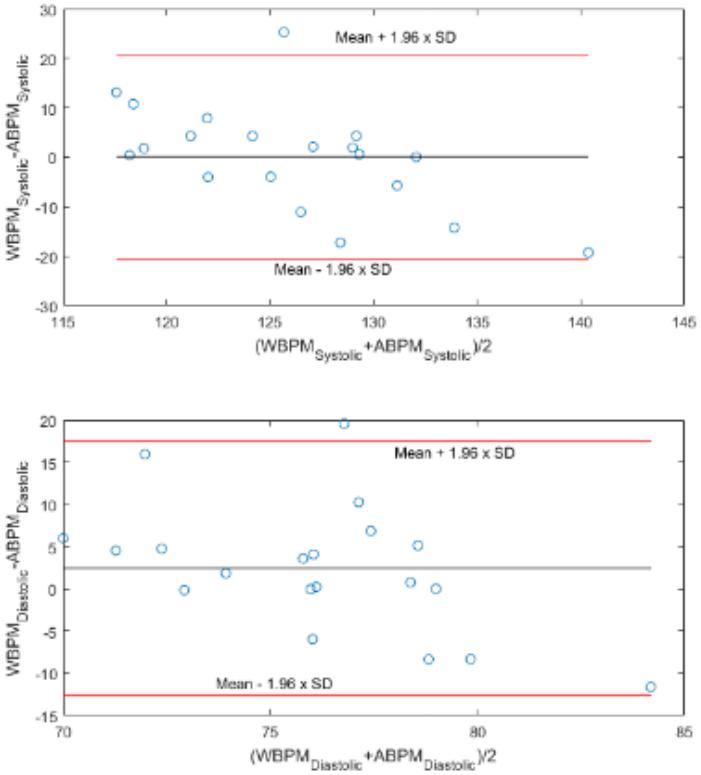


Figure 3. Bland-Altman plots between wearable and ambulatory blood pressure monitoring devices over 7 days. Measurement uncertainty during 24 hours of concurrent ambulatory blood pressure monitoring and wearable blood pressure monitoring. Black reference line represents mean bias, and red reference lines represent 95% limits of agreement. ABPM: ambulatory blood pressure monitoring; WBPM: wearable blood pressure monitoring device.



Acceptability and Barriers and Facilitators of Use

Participants were asked about the experience of using the 3 different BP devices throughout the study period. Positive and negative features of each device were clearly identified from the content analysis (Table 4). The 24-hour ambulatory device was consistently identified as being the most difficult to use and wear as it intruded into activities of daily living such as sleeping and exercising and was often uncomfortable to wear or painful during measurements. Although the home BP monitoring device was simple to use and did not interrupt daily activities, participants identified that obtaining BP measurements was reliant on them taking action and initiating a measurement. Most participants preferred the unobtrusive design and automated measurements of the wearable BP device. Although additional wearable device parameters such as heart rate and sleep quality were not formally considered in this study, participants liked being able to view these data on the device display and/or in the smartphone app. Barriers to using the wearable device included difficulty in obtaining a good fit for people with small wrists, the need for regular charging, and a motion-activated light that woke some participants during sleep.

Table 4. Advantages and disadvantages of using the study blood pressure devices.

Device	Advantages	Disadvantages
Home BP[a] monitor	P[b]05: “I thought it was so quick and easy to administer, and simple to use.”	P05: “It does rely on the participant to remember to take the readings but as long as they adhere to that...it’s quite a quick process.”
24-hour ambulatory device	P08: “I suppose the positive of that [device] is that it was automatic. Um, it’s probably the only positive.”	P14: “It’s quite stressful to wear, I think it raised my blood pressure (laughs)…because you are just there, waiting every half an hour for it to go [take measurement].”
Wearable device	P13: “If you wear the watch in the morning you’re done for the day.”; P07: “You don’t notice it at all, in terms of it collecting any measurements.”	P02: “I would question…the accuracy of the BP measurement. It just didn’t seem to match up with all the...results from either of the other ones [measurements from other devices].”; P06: “When I was asleep I must have moved my wrist. And you know how it automatically lights up? It woke me up.”

[a]BP: blood pressure.

[b]P: participant.

Although both users and health care professionals identified many advantages of the wearable BP device compared with other devices, accuracy was a consistent theme identified by both participant groups. Terms such as validity, consistency,

and reliability came up frequently in both groups, demonstrating the strong presence of this theme.

Most health care professionals described using and encouraging home BP monitoring for their patients. They acknowledged using patient-reported BP data when available, but they reported concerns regarding its reliability. When asked about the use of a wearable BP device by patients, all health care professionals expressed interest in this method as long as accuracy could be demonstrated. Health care professionals foresee that the benefits of a wearable device could include convenience (small size and regular automated measurements) and improved adherence to monitoring. Additional considerations for using wearable devices included the cost, data privacy, and use among vulnerable populations such as the elderly and those with English as a second language.

Discussion

In this study, we attempted to validate, for the first time, a wearable cuffless BP device for measuring BP continuously against a gold-standard ambulatory device and against a common home BP device. Our results suggest the wearable device compared well with the gold-standard ambulatory device over 24 hours as measurement biases were within acceptable limits, but these are not sufficient on their own to recommend wearable devices as a replacement for established ambulatory devices. In contrast, wearable BP measures differed systematically from the home device over 7 days. Given the comparability of wearable and ambulatory measures, this likely suggests the home device systematically underestimated BP. The wearable device could potentially be used for long-term BP monitoring and management if their long-term validity and reliability could be established. The wearable device was acceptable to participants and health care professionals, provided validity can be ensured.

A 7-day BP measure could provide more stable assessment of BP status than current infrequent/one-off methods, but current devices are either too cumbersome (ambulatory) or impractical (home devices) for this purpose. Our findings suggest that the 24-hour data from the wearable device were consistent with its 7-day data, which might represent true BP better than the measures from a home BP device as those readings were taken only 3 times a day compared with BP data recorded hourly in a wearable device. An important assumption is that over the 7-day period, a participant's true BP and within-participant variance could reasonably be assumed to be stable. The longer time frame allows for a sufficient number of measurements to capture the within-participant variance; however, the stability of the mean BP is less certain. The observed difference in BP between the wearable and home devices could also be because of the wearable device measuring BP during movement and daily activity, which we could not account for.

A systematic review and meta-analysis of 52 prospective studies reported that compared with usual care, self-monitoring of BP alone resulted in statistically significant improvements in SBP and DBP (−3.9 and −2.4 mm Hg, respectively) at 6 months, and that self-monitoring in combination with additional support lowered SBP from −2.1 to −8.3 mm Hg and DBP from 0.0 to −4.4 mm Hg at 12 months [18]. The unobtrusive design of the wearable device makes it well suited for assessing long-term and diurnal patterns of BP with potential prognostic significance. The wearable device used in our study is simple to set up, lightweight (23 g), waterproof, and compatible with both Android and IOS operating systems, and with a battery capacity of 80 mAh, it has a standby time of 10 days. The device can be demonstrated to the participants at the physician's clinic during consultation. As demonstrated by our qualitative data, the device is easy to use and requires minimal technical knowledge to operate, making wearable BP devices suitable for people of all ages.

A number of wearable cuffless BP devices have become available in recent years [19-23], highlighting their potential use in clinical settings. However, most of them have only been validated for point measurements in clinical consultation settings, and longer-term validity has not been assessed. The American National Standards Institute/Association for the Advancement of Medical Instrumentation and the Institute of Electrical and Electronics Engineers wearable BP standard 2014-1708 require the MAD between test and referent devices to be less than 5 mm Hg [5] for grade A classification but does not provide guidelines for continuous BP device. In this study, we used an ambulatory device as a gold standard method to assess the validity of the wearable device. As home devices are frequently recommended to patients by health care professionals, we also compared the wearable device with a home device over 7 days. Although the use of home devices is typically limited to periodic measurements in the home during static rest, the wearable device allowed assessment of BP during daily activities and sleep. This higher-frequency measurement schedule could enable assessment of single and multiday variability in BP control. Moreover, as users can view BP on the device display and/or in a smartphone app, this type of device may help participants to better understand their BP and self-initiate discussion with clinicians during consultations when required. Thus, data obtained using wearable devices over time may allow measures of monthly or annual mean BP status, which could be used to monitor treatment effectiveness, adherence, or disease progression.

A major limitation of this study was a lack of exact time synchronization between the different devices. It would be interesting to compare the patterns and validation of BP measurements between the wearable and ambulatory devices over time. However, we have estimated the daytime wearable BP with home BP measurements by identifying wearable BP records from 7 am to 9 pm as this was representative of the time range for home device measurements. Subgroup and sensitivity analyses were not planned because of the small sample size participants. The wearable device used a combination of optical sensors and software algorithms to estimate BP; however, commercially sensitivity means detailed specifications are not available; this limits comparison with another wearable device. As our study participants all had normal BP, caution is required when generalizing the results to people with high BP. Finally, it was not possible to consider potential sources of error variance such as physiological fluctuation in BP and movement artifact during physical activity.

The capability of many wearable BP devices to wirelessly interface with mobile devices and a growing number of apps/digital platforms provides a means to (1) monitor and share BP data on a long-term basis, (2) alert participants to key changes in BP, and (3) help physicians to understand treatment adherence and efficacy and make appropriate adjustments. As artificial intelligence approaches become more sophisticated, it could be possible to automate data processing and synthesis, which will help to streamline the BP monitoring workflow for time-limited physicians. Further work is needed to better understand if this wearable device can identify nighttime dipping and early morning surges in BP, which provide important clinical information about hemodynamics control. Future research into validating wearable devices in clinical populations with time-stamped data shared with patients and clinicians in a meaningful way to support clinical decision making is warranted.

In conclusion, wearable BP device measures compared well against a gold-standard ambulatory device measure. Participants found the device simple and easy to wear and use. Our findings indicate potential for this method to augment BP management, particularly by enabling long-term monitoring that could improve treatment titration and increase understanding of users' BP response during daily activity and stressors. The streamlined design and operation of wearable BP devices can offer numerous advantages compared with traditional ambulatory and home devices; however, measurement validity is a critical requirement. As this particular device did not meet established validity criteria, it cannot be recommended as a replacement for gold-standard ambulatory devices.

Acknowledgments

The authors acknowledge the support of the Institute for Physical Activity and Nutrition (IPAN) Co-Directors Alfred Deakin Professor David Crawford AM and Alfred Deakin Professor Jo Salmon for supporting this study. They acknowledge the support of their study participants; Professor Robin Daly and Dr Helen Macpherson; and IPAN, Deakin University. SMSI, SC, JR, SF and RM are funded by IPAN.

This project was funded by IPAN, School of Exercise and Nutrition Science, Faculty of Health, Deakin University.

Authors' Contributions

SMSI, CK, and RM contributed to concept and design. SMSI, JR, and RM contributed to implementation. CK, SMSI, and SC analyzed and interpreted the data. SMSI, RM, SC, JR, and SF drafted the manuscript. All authors contributed, read, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Relationships between ambulatory blood pressure monitoring and wearable blood pressure monitoring during 24 hours of concurrent monitoring.
[PDF File (Adobe PDF File), 27 KB - mhealth_v7i10e14706_app1.pdf]

Multimedia Appendix 2
Bland-Altman plots between wearable blood pressure monitoring (WBPM) and ambulatory blood pressure monitoring (ABPM) devices (24 hours). Measurement uncertainty during 24 hours of concurrent ABPM and WBPM. Black reference line represents mean bias, and red reference lines represent 95% limits of agreement.
[PDF File (Adobe PDF File), 75 KB - mhealth_v7i10e14706_app2.pdf]

Multimedia Appendix 3
Scatterplots of systolic BP versus diastolic BP for wearable and ambulatory devices. BP: blood pressure.
[PDF File (Adobe PDF File), 141 KB - mhealth_v7i10e14706_app3.pdf]

References

1. Stanaway JD, Afshin A, Gakidou E, Afshin A, Abajobir AA, Abate KH, et al. . [doi: 10.1016/S0140-6736(17)32366-8] [Medline: 30496105]
2. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. J Am Med Assoc 2017 Jan 10;317(2):165-182. [doi: 10.1001/jama.2016.19043] [Medline: 28097354]
3. Chow CK, Teo KK, Rangarajan S, Islam S, Gupta R, Avezum A, PURE (Prospective Urban Rural Epidemiology) Study Investigators. Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-,

middle-, and low-income countries. J Am Med Assoc 2013 Sep 4;310(9):959-968. [doi: 10.1001/jama.2013.184182] [Medline: 24002282]
4. Jones DW, Appel LJ, Sheps SG, Roccella EJ, Lenfant C. Measuring blood pressure accurately: new and persistent challenges. J Am Med Assoc 2003 Feb 26;289(8):1027-1030. [doi: 10.1001/jama.289.8.1027] [Medline: 12597757]
5. Parati G, Stergiou GS, Asmar R, Bilo G, de Leeuw P, Imai Y, ESH Working Group on Blood Pressure Monitoring. European Society of Hypertension practice guidelines for home blood pressure monitoring. J Hum Hypertens 2010 Dec;24(12):779-785. [doi: 10.1038/jhh.2010.54] [Medline: 20520631]
6. Imai Y, Kario K, Shimada K, Kawano Y, Hasebe N, Matsuura H, Japanese Society of Hypertension Committee for Guidelines for Self-Monitoring of Blood Pressure at Home. The Japanese Society of Hypertension guidelines for self-monitoring of blood pressure at home (second edition). Hypertens Res 2012 Aug;35(8):777-795. [doi: 10.1038/hr.2012.56] [Medline: 22863910]
7. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the eighth joint national committee (JNC 8). J Am Med Assoc 2014 Feb 5;311(5):507-520. [doi: 10.1001/jama.2013.284427] [Medline: 24352797]
8. Egan BM. Implementation of the 2017 American College of Cardiology/American Heart Association hypertension guideline. Hypertension 2019 Feb;73(2):288-290. [doi: 10.1161/HYPERTENSIONAHA.118.11712] [Medline: 30624995]
9. Parati G, Stergiou G, O'Brien E, Asmar R, Beilin L, Bilo G, European Society of Hypertension Working Group on Blood Pressure Monitoring and Cardiovascular Variability. European Society of Hypertension practice guidelines for ambulatory blood pressure monitoring. J Hypertens 2014 Jul;32(7):1359-1366. [doi: 10.1097/HJH.0000000000000221] [Medline: 24886823]
10. Islam SM, Rodgers A, Chow C. PS 15-24 a review of cuffless blood pressure measuring devices potential for use in routine clinical practice. J Hypertens 2016;34:e465. [doi: 10.1097/01.hjh.0000501232.99910.05]
11. Kim SH, Lilot M, Sidhu KS, Rinehart J, Yu Z, Canales C, et al. Accuracy and precision of continuous noninvasive arterial pressure monitoring compared with invasive arterial pressure: a systematic review and meta-analysis. Anesthesiology 2014 May;120(5):1080-1097. [doi: 10.1097/ALN.0000000000000226] [Medline: 24637618]
12. Alpert BS, Quinn DE, Friedman BA. A review of the latest guidelines for NIBP device validation. Blood Press Monit 2013 Dec;18(6):297-302. [doi: 10.1097/MBP.0000000000000007] [Medline: 24192842]
13. Mancia G, Fagard R, Narkiewicz K, Redon J, Zanchetti A, Böhm M, et al. 2013 ESH/ESC guidelines for the management of arterial hypertension. Blood Press 2013 Aug;22(4):193-278. [doi: 10.3109/08037051.2013.812549] [Medline: 23777479]
14. Palatini P, Frigo G, Bertolo O, Roman E, da Cortà R, Winnicki M. Validation of the A&D TM-2430 device for ambulatory blood pressure monitoring and evaluation of performance according to subjects' characteristics. Blood Press Monit 1998 Aug;3(4):255-260. [Medline: 10212363]
15. 1708-2014 - IEEE Standard for Wearable Cuffless Blood Pressure Measuring Devices. New York, USA: IEEE; 2014.
16. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998 Oct;26(4):217-238. [doi: 10.2165/00007256-199826040-00002] [Medline: 9820922]
17. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
18. Uhlig K, Patel K, Ip S, Kitsios GD, Balk EM. Self-measured blood pressure monitoring in the management of hypertension: a systematic review and meta-analysis. Ann Intern Med 2013 Aug 6;159(3):185-194. [doi: 10.7326/0003-4819-159-3-201308060-00008] [Medline: 23922064]
19. Garcia-Ortiz L, Recio-Rodríguez JI, Canales-Reina JJ, Cabrejas-Sánchez A, Gomez-Arranz A, Magdalena-Belio JF, EVIDENT Group. Comparison of two measuring instruments, B-pro and SphygmoCor system as reference, to evaluate central systolic blood pressure and radial augmentation index. Hypertens Res 2012 Jun;35(6):617-623. [doi: 10.1038/hr.2012.3] [Medline: 22297480]
20. Zheng YL, Yan BP, Zhang YT, Poon CC. An armband wearable device for overnight and cuff-less blood pressure measurement. IEEE Trans Biomed Eng 2014 Jul;61(7):2179-2186. [doi: 10.1109/TBME.2014.2318779] [Medline: 24760899]
21. Zheng Y, Poon CC, Yan BP, Lau JY. Pulse arrival time based cuff-less and 24-h wearable blood pressure monitoring and its diagnostic value in hypertension. J Med Syst 2016 Sep;40(9):195. [doi: 10.1007/s10916-016-0558-6] [Medline: 27447469]
22. Schoot TS, Weenk M, van de Belt TH, Engelen LJ, van Goor H, Bredie SJ. A new cuffless device for measuring blood pressure: a real-life validation study. J Med Internet Res 2016 May 5;18(5):e85 [FREE Full text] [doi: 10.2196/jmir.5414] [Medline: 27150527]
23. Peter L, Noury N, Cerny M. A review of methods for non-invasive and continuous blood pressure monitoring: pulse transit time method is promising? IRBM 2014 Oct;35(5):271-282. [doi: 10.1016/j.irbm.2014.07.002]

Abbreviations

BP: blood pressure
DBP: diastolic blood pressure
IPAN: Institute for Physical Activity and Nutrition

LOA: limits of agreement
MAD: mean absolute difference
MAPD: mean absolute percentage difference
SBP: systolic blood pressure

Edited by G Eysenbach; submitted 13.05.19; peer-reviewed by MA Moni, P Banik; comments to author 16.06.19; revised version received 27.06.19; accepted 28.06.19; published 14.09.19.

Please cite as:
Islam SMS, Cartledge S, Karmakar C, Rawstorn JC, Fraser SF, Chow C, Maddison R
Validation and Acceptability of a Cuffless Wrist-Worn Wearable Blood Pressure Monitoring Device Among Users and Health Care Professionals: Mixed Methods Study
JMIR Mhealth Uhealth 2019;7(10):e14706
URL: https://mhealth.jmir.org/2019/10/e14706
doi:10.2196/14706
PMID:31628788

Original Paper

Heart Rate Measures From Wrist-Worn Activity Trackers in a Laboratory and Free-Living Setting: Validation Study

Andre Matthias Müller[1,2], PhD; Nan Xin Wang[1], MSc; Jiali Yao[3], MSc; Chuen Seng Tan[3,4], PhD; Ivan Cherh Chiet Low[5], PhD; Nicole Lim[6], BBA; Jeremy Tan[6], BEng; Agnes Tan[6], BSc; Falk Müller-Riemenschneider[1,4,7], MBBS, MPH, MD

[1]Health Systems & Behavioral Sciences, Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

[2]Centre for Sport & Exercise Sciences, University of Malaya, Kuala Lumpur, Malaysia

[3]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

[4]Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore

[5]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[6]Health Promotion Board Singapore, Singapore, Singapore

[7]Institute for Social Medicine, Epidemiology and Health Economics, Charite University Medical Centre Berlin, Berlin, Germany

Corresponding Author:
Andre Matthias Müller, PhD
Health Systems & Behavioral Sciences
Saw Swee Hock School of Public Health
National University of Singapore and National University Health System
Tahir Foundation Building
12 Science Drive 2, #09-01C
Singapore, 117549
Singapore
Phone: 65 82997548
Email: ephamm@nus.edu.sg

Abstract

Background: Wrist-worn activity trackers are popular, and an increasing number of these devices are equipped with heart rate (HR) measurement capabilities. However, the validity of HR data obtained from such trackers has not been thoroughly assessed outside the laboratory setting.

Objective: This study aimed to investigate the validity of HR measures of a high-cost consumer-based tracker (Polar A370) and a low-cost tracker (Tempo HR) in the laboratory and free-living settings.

Methods: Participants underwent a laboratory-based cycling protocol while wearing the two trackers and the chest-strapped Polar H10, which acted as criterion. Participants also wore the devices throughout the waking hours of the following day during which they were required to conduct at least one 10-min bout of moderate-to-vigorous physical activity (MVPA) to ensure variability in the HR signal. We extracted 10-second values from all devices and time-matched HR data from the trackers with those from the Polar H10. We calculated intraclass correlation coefficients (ICCs), mean absolute errors, and mean absolute percentage errors (MAPEs) between the criterion and the trackers. We constructed decile plots that compared HR data from Tempo HR and Polar A370 with criterion measures across intensity deciles. We investigated how many HR data points within the MVPA zone (≥64% of maximum HR) were detected by the trackers.

Results: Of the 57 people screened, 55 joined the study (mean age 30.5 [SD 9.8] years). Tempo HR showed moderate agreement and large errors (laboratory: ICC 0.51 and MAPE 13.00%; free-living: ICC 0.71 and MAPE 10.20%). Polar A370 showed moderate-to-strong agreement and small errors (laboratory: ICC 0.73 and MAPE 6.40%; free-living: ICC 0.83 and MAPE 7.10%). Decile plots indicated increasing differences between Tempo HR and the criterion as HRs increased. Such trend was less pronounced when considering the Polar A370 HR data. Tempo HR identified 62.13% (1872/3013) and 54.27% (5717/10,535) of all MVPA time points in the laboratory phase and free-living phase, respectively. Polar A370 detected 81.09% (2273/2803) and 83.55% (9323/11,158) of all MVPA time points in the laboratory phase and free-living phase, respectively.

Conclusions: HR data from the examined wrist-worn trackers were reasonably accurate in both the settings, with the Polar A370 showing stronger agreement with the Polar H10 and smaller errors. Inaccuracies increased with increasing HRs; this was pronounced for Tempo HR.

(JMIR Mhealth Uhealth 2019;7(10):e14120) doi:10.2196/14120

KEYWORDS

eHealth; mHealth; wearable; exercise; measurement; fitness; public health; quantified self

Introduction

The scientific evidence on the health and well-being benefits of physical activity (PA) is overwhelming, and, as such, increasing activity levels is a core public health target [1,2]. The greatest benefits are attained when engaging in regular moderate-to-vigorous PA (MVPA). For example, positive effects of MVPA have been shown in the domains of mortality risk [3,4] as well as physical and psychological health and well-being [1,5-7].

The PA research landscape could broadly be divided into 2 core facets—PA surveillance and PA promotion [1,2]. Key to progress in both is the accurate measurement of PA. In this regard, questionnaires that are prone to recall and social desirability bias [8] are increasingly being complemented by instruments that measure PA objectively. Wearable research-grade devices such as accelerometers are commonly used, which have improved the validity of PA estimates [9,10]. Unfortunately, conducting population-wide studies with accelerometers is difficult because of high costs and participant burden.

To this end, the soaring availability and use of commercial wrist-worn activity tracking devices are increasingly being harnessed by PA researchers who are keen to use them for large-scale surveillance and intervention studies [11-13]. Sensor technologies inbuilt in these trackers allow for the convenient collection of various types of data [14]. As such, observational research and PA monitoring in intervention studies might soon primarily rely on data collected through devices that were not developed for research purposes. However, to make adequate use of wrist-worn tracker data, the validity of such data needs to be established [15]. There are numerous validation studies that have been conducted in recent years, and most of the studies focused on the accuracy of accelerometer-based metrics (eg, step counts) that are available from generation 1 activity trackers [16-18].

In addition to measurements of accelerometer-based metrics, many newer wrist-worn trackers are equipped with capabilities to collect data on physiological measures such as heart rate (HR) [14]. Estimating HR is enabled through photoplethysmography (PPG), a technology that consists of light-emitting diodes and photodetectors. With this, volume changes in the pulsatile component in the microvascular bed of arterial blood can be captured through reflection of the emitted light through the tissue [19]. Algorithms are then applied to estimate HR from PPG information. The 2 key advantages to measuring HR instead of, or in addition to, other metrics are the capture of nonweight-bearing activities (eg, cycling) and the ability to ascertain PA intensity, which is important for MVPA monitoring.

Validating HR data from wrist-worn trackers in healthy individuals is a recent endeavor [14]. Despite the increasing research activity, there is currently little uniformity in the technologies used and the conditions in which studies have been conducted. For example, although most researchers assessed the accuracy of tracker-based HR data during cycling or treadmill exercises [20-36], protocols varied widely (eg, different speeds and varying durations). Some of these studies also examined tracker accuracy during chores [31], outdoor activities [33], and resistance exercises [32,35,36], and 1 research team was solely interested in HR data accuracy during sedentary time [37]. As such, drawing firm and generalizable conclusions about tracker accuracy in terms of HR data is difficult, and interested readers are advised to consult studies that assessed the accuracy of specific trackers during specific activities (eg, accuracy of the various Fitbit devices (Fitbit, Inc) during cycling).

What the above-mentioned studies have in common is that they were conducted in a controlled laboratory setting. However, there are differences between a controlled and less controlled environment, and collecting data in both environments is warranted to disentangle such differences and increase ecological relevance of findings. To our knowledge, there are only 2 free-living studies. One research team merely collected HR data during common daily activities over a few hours [38], whereas the other included only 1 participant [39]. This is unsatisfying because wrist-worn trackers are meant to accurately capture HRs in different PA intensity zones (eg, light PA and MVPA) throughout the day and in different people. In addition, sample sizes were mostly small, with only few studies having more than 50 participants [22,28,40]. Finally, HR validation studies have so far only included devices that are rather expensive (mainly devices from Fitbit; 40 of 61 validation studies) [14]. With this, many people who might benefit from trackers will not be able to afford them. Less expensive trackers are readily available, but they are rarely tested. If these more affordable devices are reasonably accurate, large-scale studies and population-based health promotion campaigns using such activity trackers could become commonplace.

This study aimed to examine the validity of HR data from 2 wrist-worn HR trackers, the Tempo HR, a low-cost device used for a national PA promotion campaign in Singapore, and the Polar A370, a consumer-based fitness and activity tracking device, in laboratory and free-living settings. Both these trackers have not been assessed previously.

Methods

Design

We conducted a 2-phased validation study with all participants: laboratory phase and free-living phase. The study procedures were approved by the institutional review board of the National University of Singapore (NUS IRB: S-18-026), and written-informed consent was obtained from all participants before study enrolment. Data collection took place between March and May 2018.

Participants

We applied multiple recruitment strategies to ensure a sample with varied characteristics. Students and staff were recruited through a post on the university's Web-based learning system blackboard and word-of-mouth. Participants from the general public were recruited through emails sent to participants of the National Steps Challenge (NSC), a national PA promotion campaign rolled out by the Health Promotion Board (HPB), Singapore, yearly for 6 months (October to April).

Interested people were assessed for eligibility during an initial screening call and during the laboratory visit. The following inclusion criteria were applied: reasonably physically active English-literate men and women aged between 21 and 50 years with a body mass index (BMI) of at least 18.5 kg/m^2; absence of physical disabilities or illness that would restrict moderate PA as assessed with the Physical Activity Readiness Questionnaire [41]; ownership of a mobile phone that supports HPB's Healthy365 app, which was needed for data retrieval from the Tempo HR tracker, and that is compatible with HPB's activity trackers; and willingness to use the personal mobile phone. Participants were instructed to abstain from caffeine for 12 hours and from food for 2 hours before the first study center visit.

Procedures: Laboratory Phase

During the first visit, we collected sociodemographic information and measured height and weight with a SECA stadiometer (SECA GmbH). Following this, participants were fitted with 3 HR monitoring devices. We used the chest-strapped Polar H10 HR monitor (Polar Electro Oy) as our criterion device. Concurrent validity of similar Polar devices against echocardiogram (ECG) is well established [42]. The device was placed below the chest muscles. It transmitted real-time HR data to a wristwatch via Bluetooth. Our 2 wrist-worn HR trackers were the trackers used for the NSC (Tempo HR, J-style, TEMPO) and the Polar A370 (Polar Electro Oy). The Tempo HR is a low-cost activity tracker that measures steps, distance, calories burnt, and HR. Data from this tracker were transferred to the participants' Healthy365 app, downloaded by HPB staff at the backend before it was shared with the researchers. Those recruited through HPB were already in possession of the Tempo HR tracker. The Polar A370 is a commercial activity tracker that allows monitoring of steps, distances, pace, global positioning system location, calories burnt, and HR. Data were transferred to the associated Polar Flow app and downloaded to the computer. Devices were worn snugly on opposite wrists (Tempo HR: left and Polar A370: right, during both the phases).

Resting HR following at least 5 min of continuous sitting was measured before the cycling protocol.

Participants were requested to go through an incremental cycling protocol of 20 min on a stationary exercise bicycle (Monark 894E). The protocol consisted of four 5-min stages, and participants were required to cycle at an intensity corresponding to their designated HR zones for each stage (45%, 55%, 65%, and 75% of maximum HR [HR_{max}]; ±10 beats per minute [bpm]) [22]. HR_{max} was calculated according to the common formula 220−age in years [43]. During the cycling program, researchers monitored adherence to the HR zones, provided verbal encouragement if necessary, and recorded perceived exertion at midpoint of each stage using the well-established 15-point visual Borg scale [44]. Following the cycling program, participants' recovery HR was monitored for 5 min.

Procedures: Free-Living Phase

After completing the cycling protocol, participants were introduced to the procedures of the free-living phase. In addition to the devices used in the laboratory phase, we provided participants with an ActiGraph wGT3X+BT accelerometer (ActiGraph) to collect HR data from the Polar H10 chest strap via Bluetooth. The small tamper-proof device was attached with a belt to the right side of the hip. We also provided an instruction sheet detailing adequate wear.

Participants were instructed to wear the devices during waking hours of the following day (after getting up in the morning until bedtime at night) and only remove them during water-based activities. In addition, we requested that participants engage in at least one 10-min bout of MVPA during the day to capture a wide range of HR signals. Finally, participants were provided with a device-wear log to record their wear and nonwear as well as their MVPA session(s). Participants returned to the laboratory a few days later to return the study devices and transfer HR data of the Tempo HR to the Healthy365 app.

Data Acquisition and Synchronization

The sampling frequencies of the Tempo HR, Polar A370, and Polar H10 chest strap were 0.1 Hz, 1 Hz, and 1 Hz, respectively. As such, HR data were collected every second by the Polar devices and every 10 seconds by the Tempo HR (a sample of the raw data is provided in Multimedia Appendix 1). All devices provided time-stamped HR data based on the Network Time Protocol (GMT plus 8 hours). This allowed for time matching of data. For our analyses, we extracted the 10-second values from all 3 devices and time matched the nonzero HR data from Tempo HR and Polar A370 with those from Polar H10. The following data inclusion criteria were applied for the 2 phases separately: availability of at least 10 min of time-matched data for the laboratory phase and availability of at least 180 min of time-matched data for the free-living phase.

Statistical Analysis

We summarized participants' characteristics descriptively using mean and SD for continuous variables and number and percentage for categorical variables.

We calculated the intraclass correlation coefficients (ICCs) using mixed effects models to assess the absolute agreement

between the criterion (Polar H10) and the other trackers (Tempo HR and Polar A370) in the laboratory phase and free-living phase. The strength of the ICC was interpreted as weak (<0.50), moderate (≥0.50 to 0.74), strong (≥0.75 to 0.89), and very strong (≥0.90) [45]. To facilitate visual inspection, we created scatterplots of HRs between devices with all participants and similarly for each participant (not shown).

We then calculated mean absolute errors (MAEs) and mean absolute percentage errors (MAPE; absolute error/criterion×100) between the criterion (Polar H10) and, both, the Tempo HR and the Polar A370 trackers, to gauge overall measurement error. As highlighted in a recent study, there is no clear cutoff for what level of error would indicate adequate validity between measures [32]. After considering the available options and similar to the authors of a previous study, we adopted a cutoff of 10% to judge validity [46], a cutoff that also coincides with the one suggested by the Association for the Advancement of Medical Instrumentation in their document on the validity of HR measurement devices [47]. Bland-Altman (BA) plots with limits of agreement (LoA) set at 95% were used to visualize agreement and proportional bias.

Moreover, we ranked the 10-second HR time points derived from the Polar H10 and divided them into deciles. As such, decile 1 contained the lowest 10% of all HR and decile 10 contained the highest 10% of all HR. We then time matched these HR deciles with HR data from the Tempo HR and Polar A370. We constructed the box plots to compare the HR data from the Tempo HR and the Polar A370 with the Polar H10 measures across the deciles.

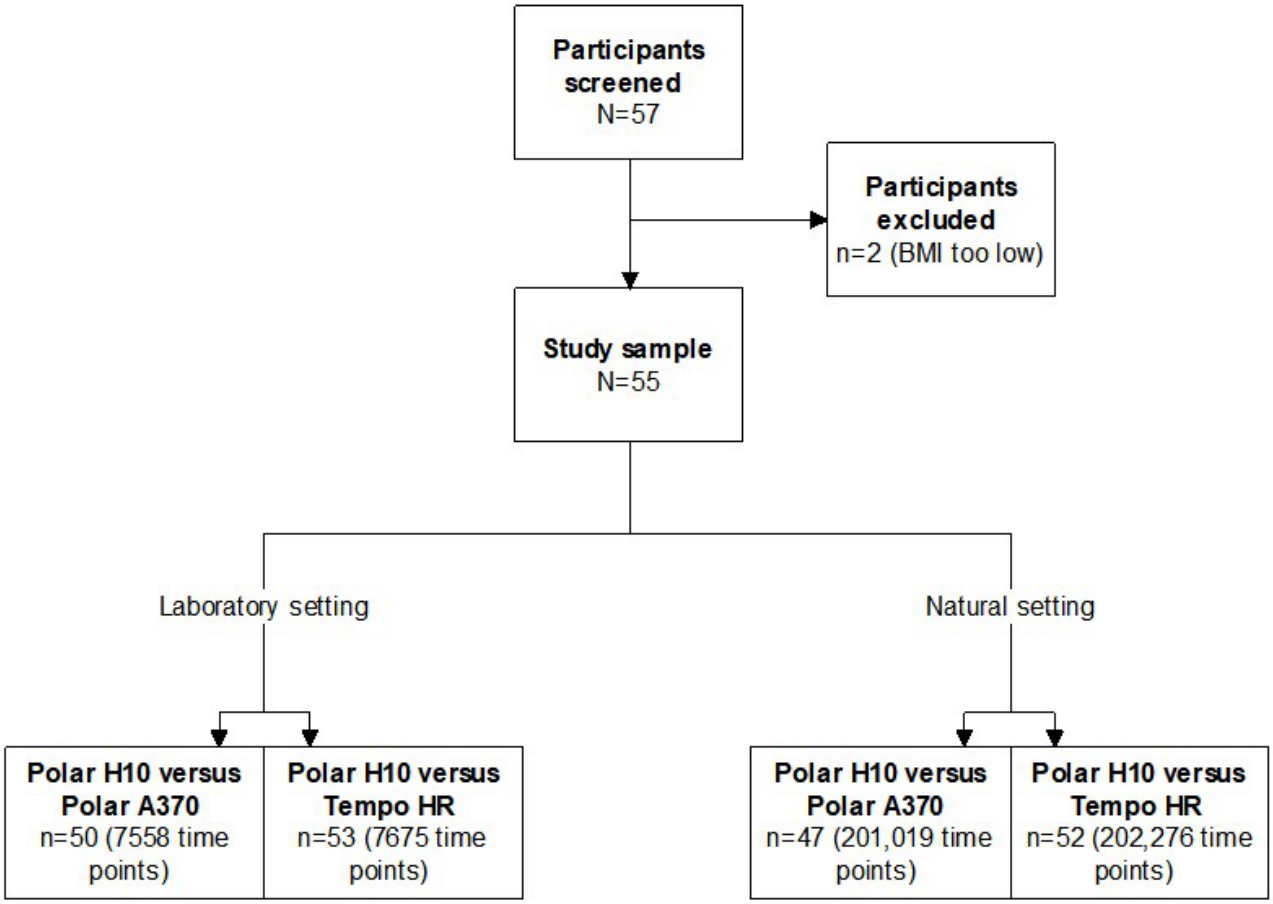
Finally, we constructed 2×2 tables to estimate the sensitivity and specificity of the 2 trackers for identifying the different HR zones based on the Polar H10 (<64% HR_{max} and ≥64% HR_{max}). The cutoff of 64% HR_{max} was chosen because it is the updated cutoff [48] of the earlier 50% HR_{max} cutoff [49]. The more recent cutoff has since been endorsed by the American College of Sports Medicine [50]. All statistical analyses were conducted using R (version 3.4.2).

Results

Study Participants

Of the 57 people screened, 55 were eligible and joined the study (mean age 30.5 [SD 9.8] years), with 26 being female (47%), 36 with normal weight (65%; BMI <23 kg/m^2), and 39 with Chinese ethnicity (71%). Due to the unavailability of some HR data, few participants were excluded from some analyses. Figure 1 depicts the analysis flow, which also indicates data availability. During the free-living phase, and after excluding data points with zero measures, mean wear time of the Tempo HR, Polar A370, and Polar H10 was 12.2 (SD 2.6) hours, 12.8 (SD 2.7) hours, and 11.7 (SD 3.1) hours, respectively.

Figure 1. Data analysis flow showing participants in analysis (n) and number of matched heart rate time points. BMI: body mass index; HR: heart rate.



Overall Agreement

Laboratory Phase

In the laboratory phase, the HR data from the Tempo HR showed a moderate ICC (0.51; 95% CI 0.38 to 0.60) with the data from Polar H10. With a MAE of 15.1 bpm (95% CI 14.6 to 15.5 bpm) and an MAPE of 13.0%, the measurement error was somewhat large. Polar A370 data also had a moderate but stronger ICC with the Polar H10 (0.73; 95% CI 0.66 to 0.78). Measurement errors were small with a MAE of 7.3 bpm (95% CI 7.0 to 7.7 bpm) and an MAPE of 6.4%. On average, both the devices underestimated HR: Tempo HR by 9.7 bpm (95% CI −10.2 to −9.2 bpm) and Polar A370 by 5.7 bpm (95% CI −6.1 to −5.3 bpm).

Figure 2 shows the BA plot, and Figure 3 shows the HR decile plot between the Tempo HR and the Polar H10. These plots showcase 3 trends: HR tends to be underestimated by the Tempo HR across the range of HR values; the increase in HR is accompanied by an increasing difference between the Tempo HR and Polar H10 HR data; and the variability of the HR data from the Tempo HR increases with increasing HR and the variability is especially pronounced at the higher HR deciles (Figure 3).

Figure 2. Laboratory phase: Bland-Altman plot between the heart rate data from the Polar H10 and the Tempo HR. Light blue dotted lines show the limits of agreement, and the dark blue dotted line shows the mean of the difference. HR: heart rate.

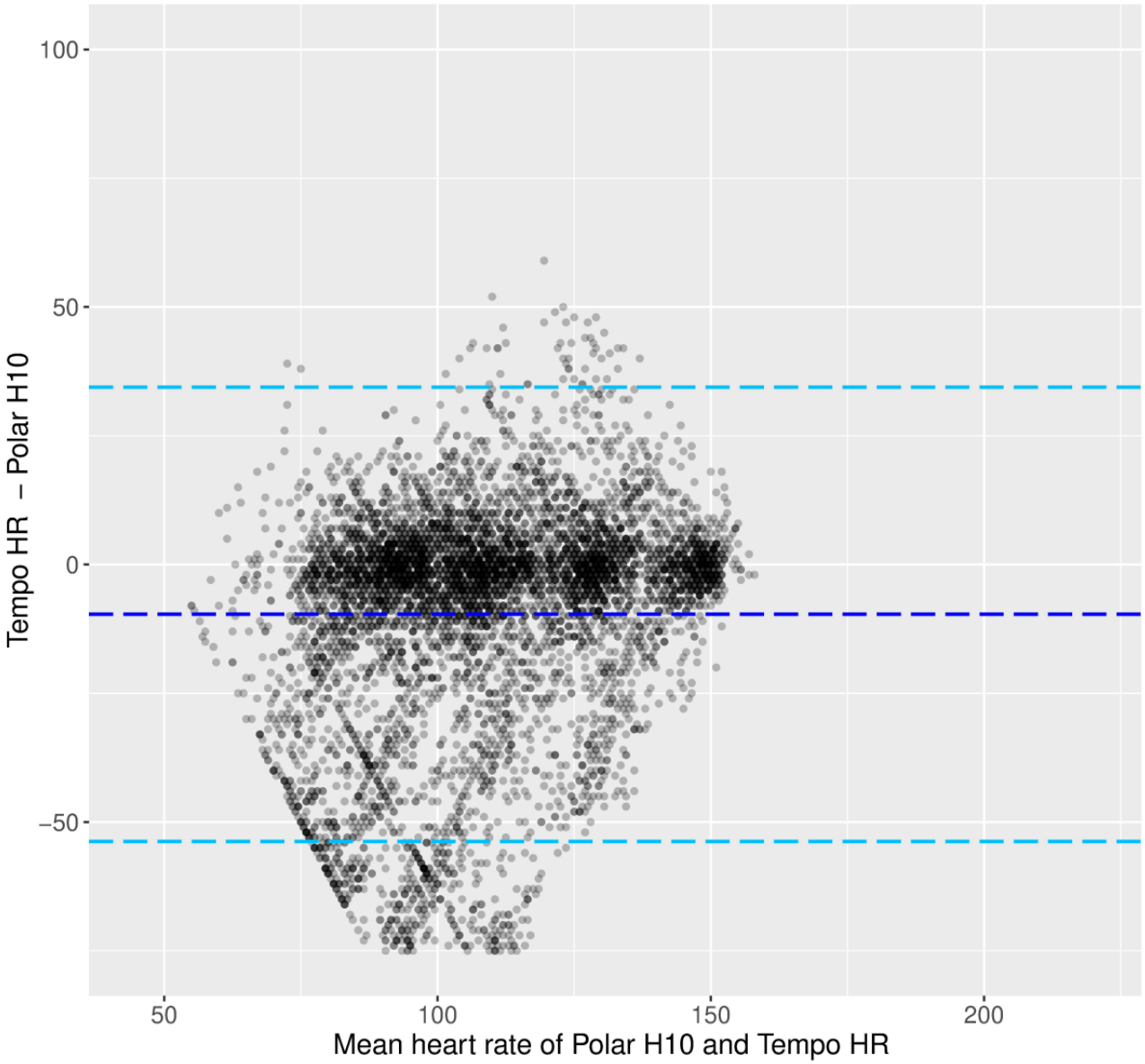
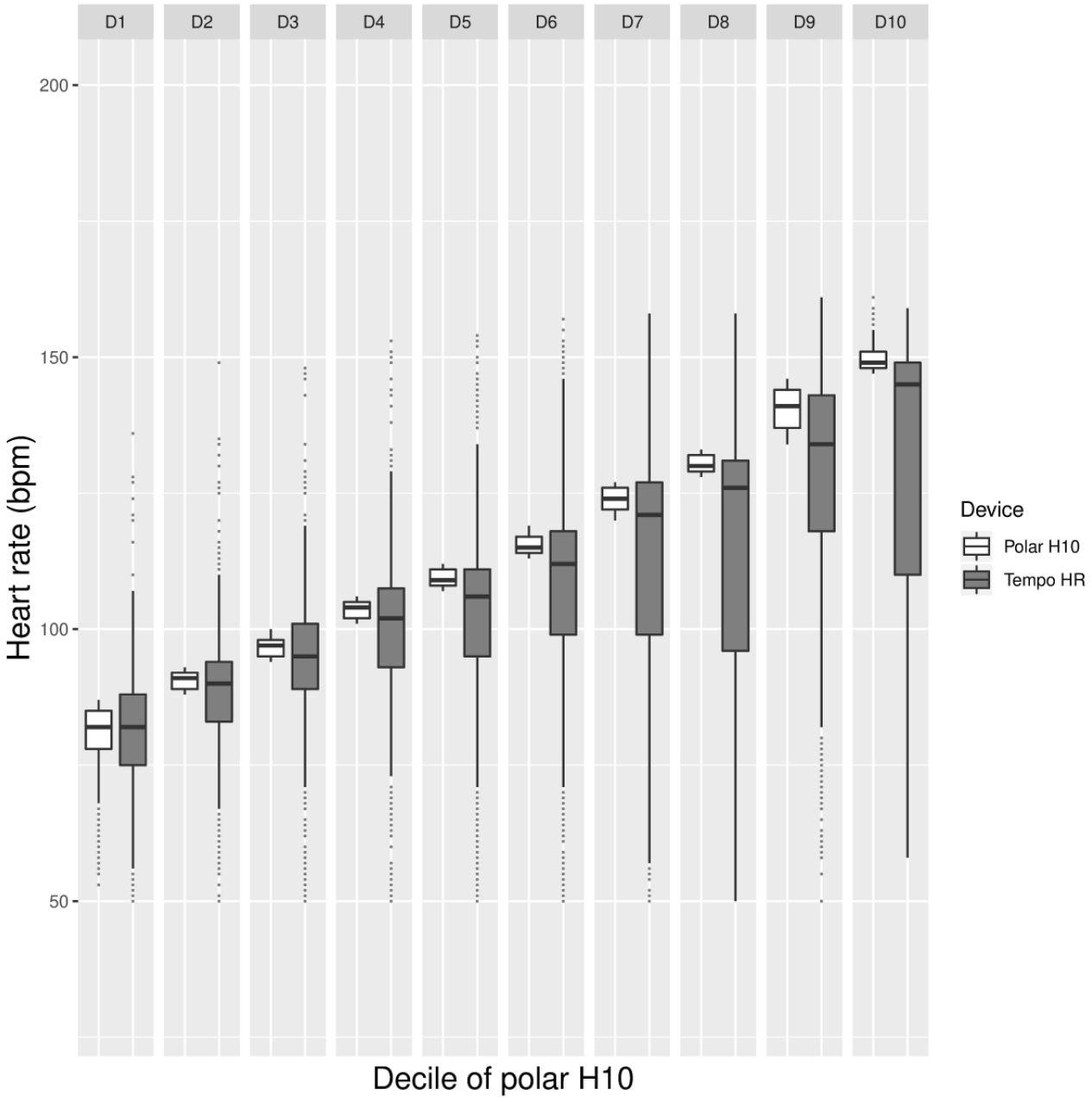


Figure 3. Laboratory phase: box plot providing by-decile comparisons of mean Polar H10 (white) and Tempo HR HR data (gray). HR: heart rate; bpm: beats per minute.



As can be seen in Figures 4 and 5, the trends described above are less pronounced when considering HR data from the Polar A370 tracker. First, it can be seen that the underestimation of HR is occurring across HR values (Figure 4). Second, the decile plot does not indicate a marked change in the difference between the data from the Polar A370 and the Polar H10 across HRs. Third, the variability of Polar A370 does not increase markedly with increasing HR (Figure 5).

Figure 4. Laboratory phase: Bland-Altman plot between the heart rate data from the Polar H10 and the Polar A370. Light blue dotted lines show the limits of agreement, and the dark blue dotted line shows the mean of the difference. HR: heart rate.

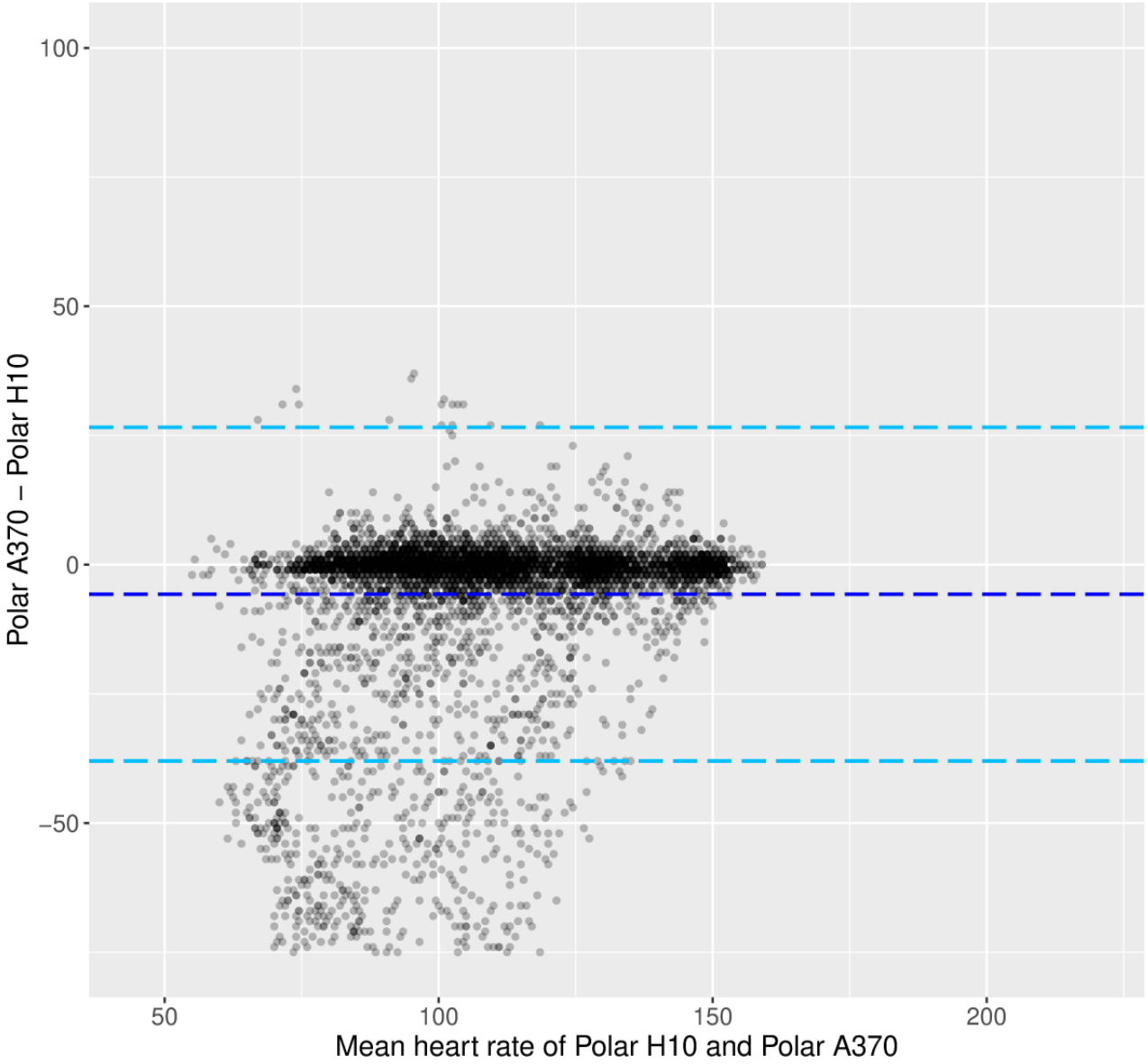
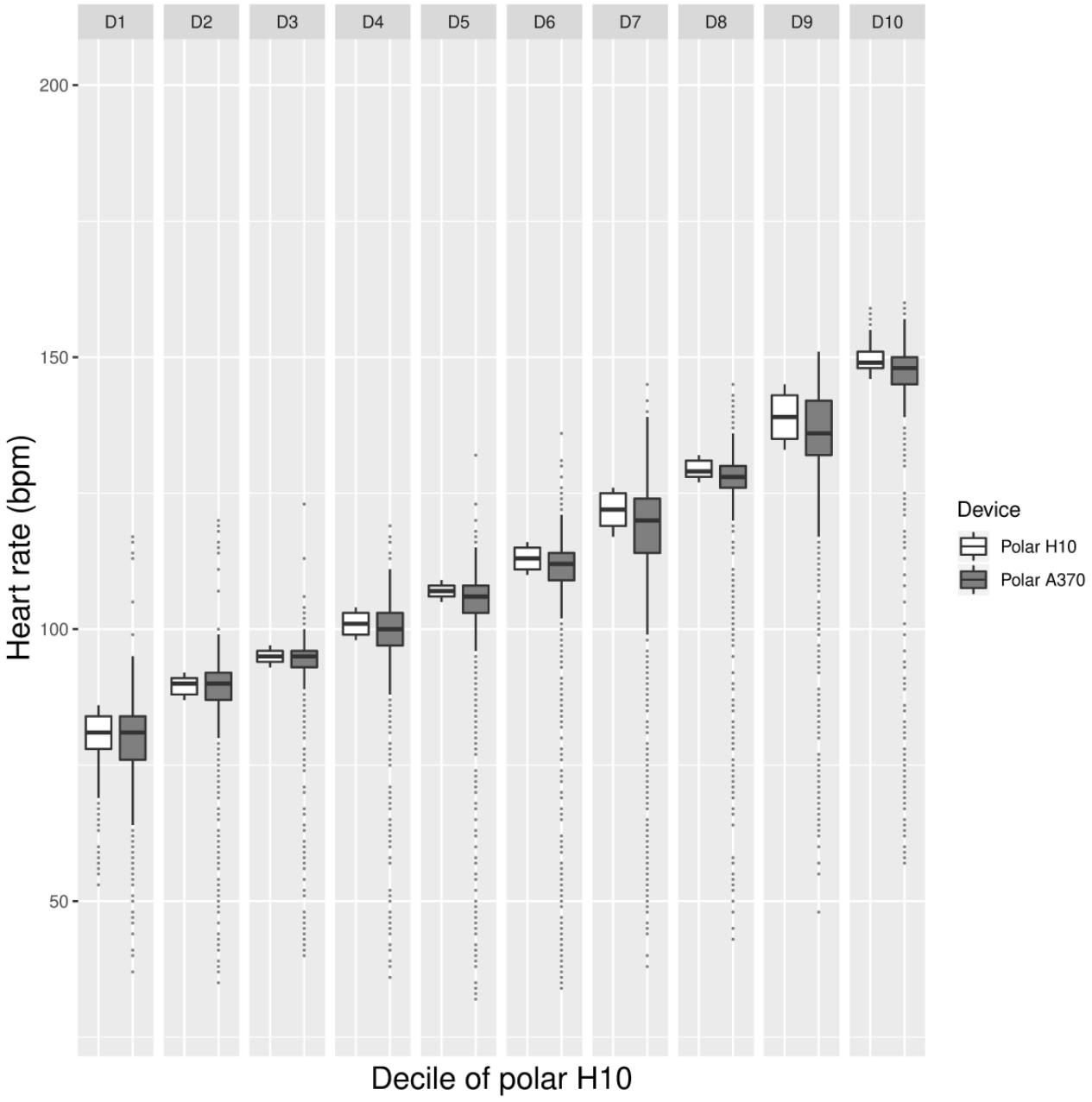


Figure 5. Laboratory phase: box plot providing by-decile comparisons of mean Polar H10 (white) and Polar A370 HR data (gray). HR: heart rate; bpm: beats per minute.



Free-Living Phase

The ICC between the Polar H10 and the Tempo HR data was moderate in the free-living phase (0.71; 95% CI 0.70 to 0.71). Errors were smaller compared with the laboratory phase with a MAE of 8.7 bpm (95% CI 8.7 to 8.8 bpm) and an MAPE of 10.2%. For the Polar A370, the ICC between the Polar H10 and the Polar A370 tracker data was strong (0.83; 95% CI 0.79 to 0.87). Errors were similar compared with the ones in the laboratory phase with a MAE of 5.9 bpm (95% CI 5.8 to 5.9 bpm) and an MAPE of 7.1%. In contrast to the results from the laboratory phase, both the devices overestimated HR slightly (Tempo HR 0.4 bpm; 95% CI 0.3 to 0.5 bpm and Polar A370 3.4 bpm; 95% CI 3.3 to 3.4 bpm).

The BA plot in Figure 6 depicts the potential occurrences of overestimation and underestimation of HR measures from the Tempo HR across HR values. Although no clear trend can be established, it appears that HR overestimation is more common. As shown in Figure 7, overestimation tends to occur at lower HRs, whereas underestimation happens more frequently at higher HRs. In addition, the decile plot shows that the HR difference between the Polar H10 and the Tempo HR is minimal until decile 8 where it begins to increase markedly. At decile 10, the difference is substantial. In addition, Tempo HR data vary to a similar degree until decile 10, where the variability is high.

Figure 8 shows the BA plot, and Figure 9 shows the HR decile plot between the Polar A370 and the Polar H10. Both plots indicate that the Polar A370 appears to overestimate HR at lower HRs (below decile 9). However, the decile plot shows that the overall difference between the criterion and the Polar

A370 is not substantial throughout. As for the Tempo HR, data variability is high only in decile 10.

Figure 6. Free-living phase: Bland-Altman plot between the heart rate data from the Polar H10 and the Tempo HR. Light blue dotted lines show the limits of agreement, and the dark blue dotted line shows the mean of the difference. HR: heart rate.

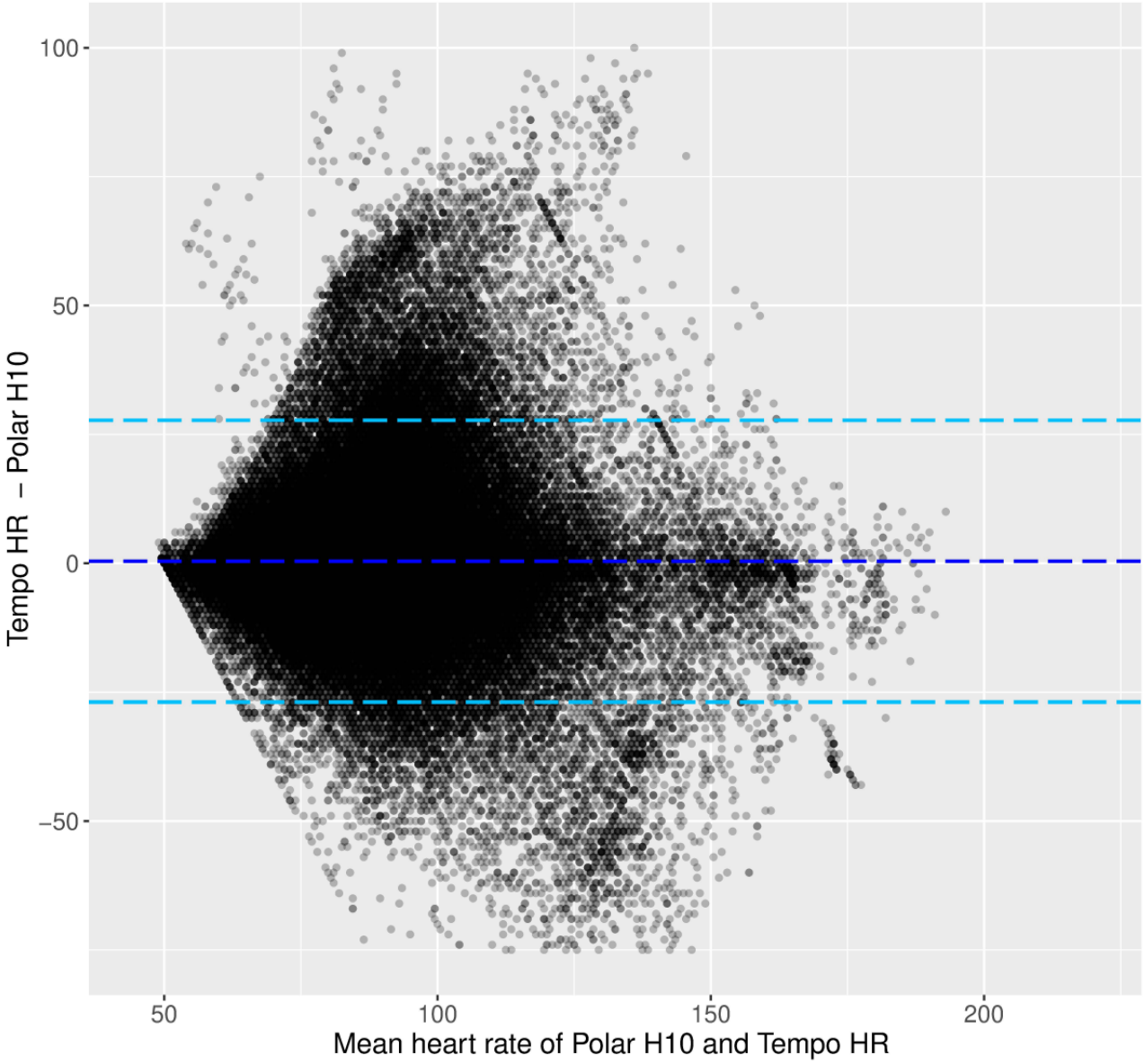


Figure 7. Free-living phase: box plot providing by-decile comparisons of mean Polar H10 (white) and Tempo HR HR data (gray). HR: heart rate; bpm: beats per minute.

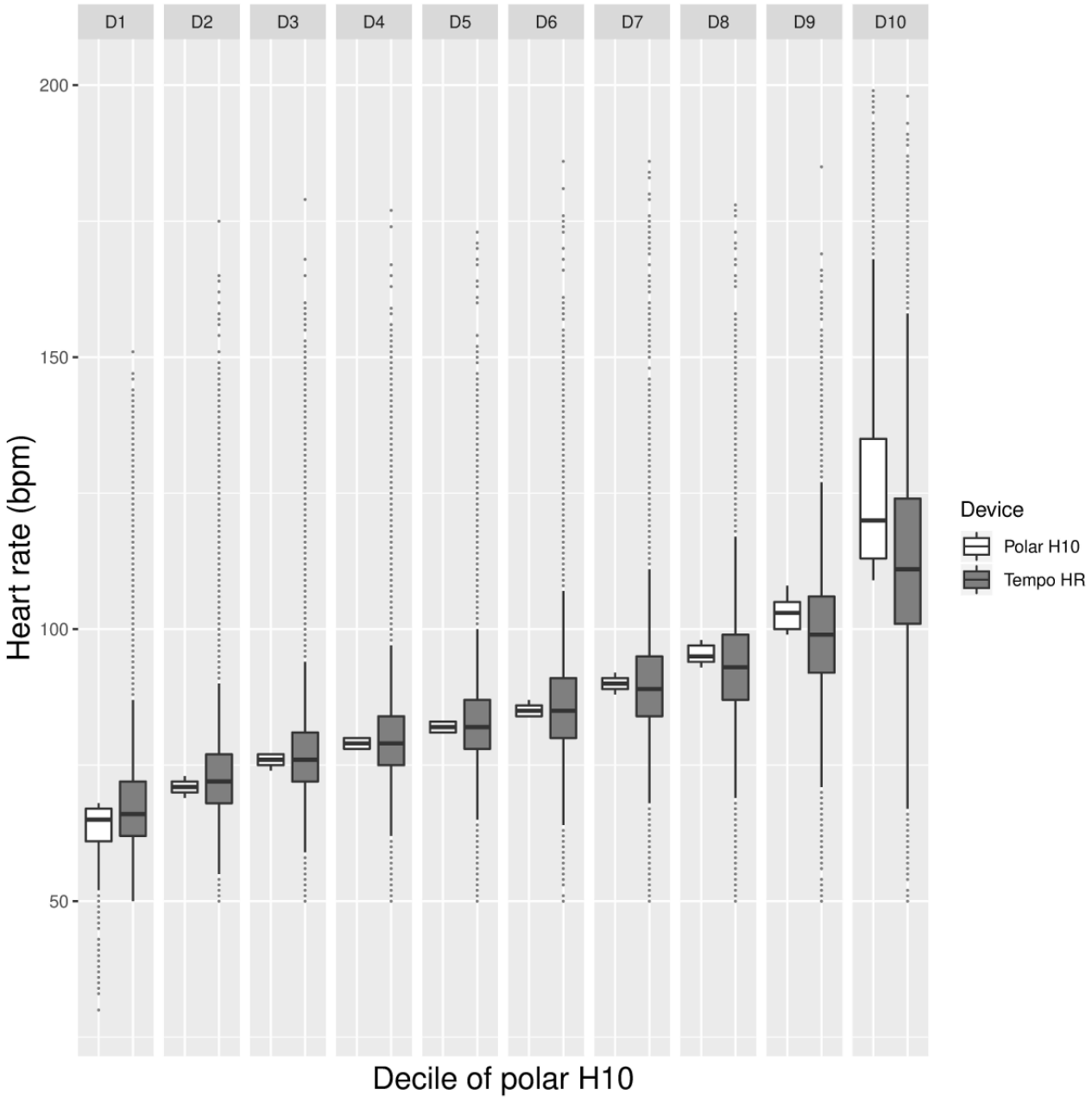


Figure 8. Free-living phase: Bland-Altman plot between the heart rate data from the Polar H10 and the Polar A370. Light blue dotted lines show the limits of agreement, and the dark blue dotted line shows the mean of the difference. HR: heart rate.

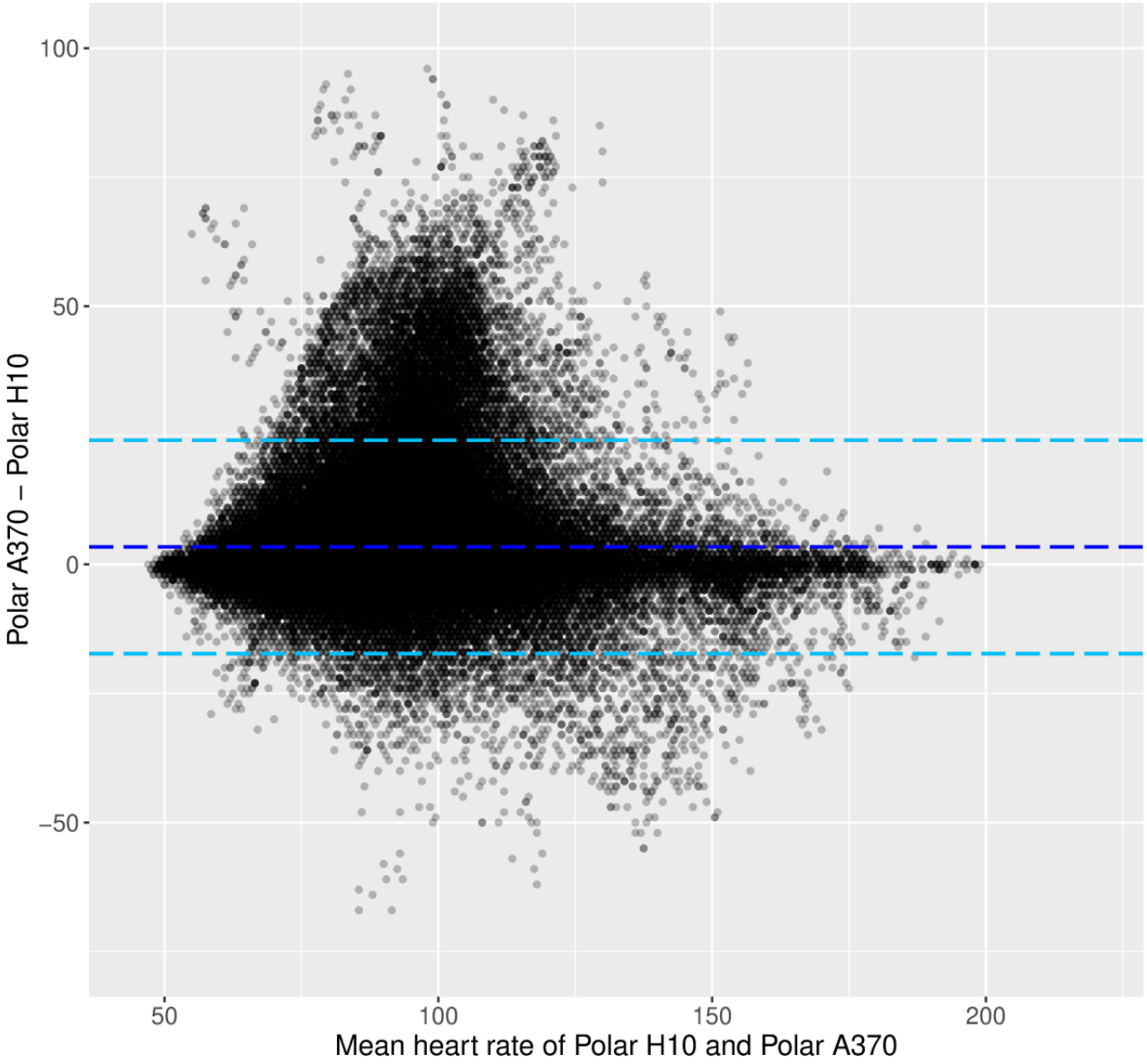
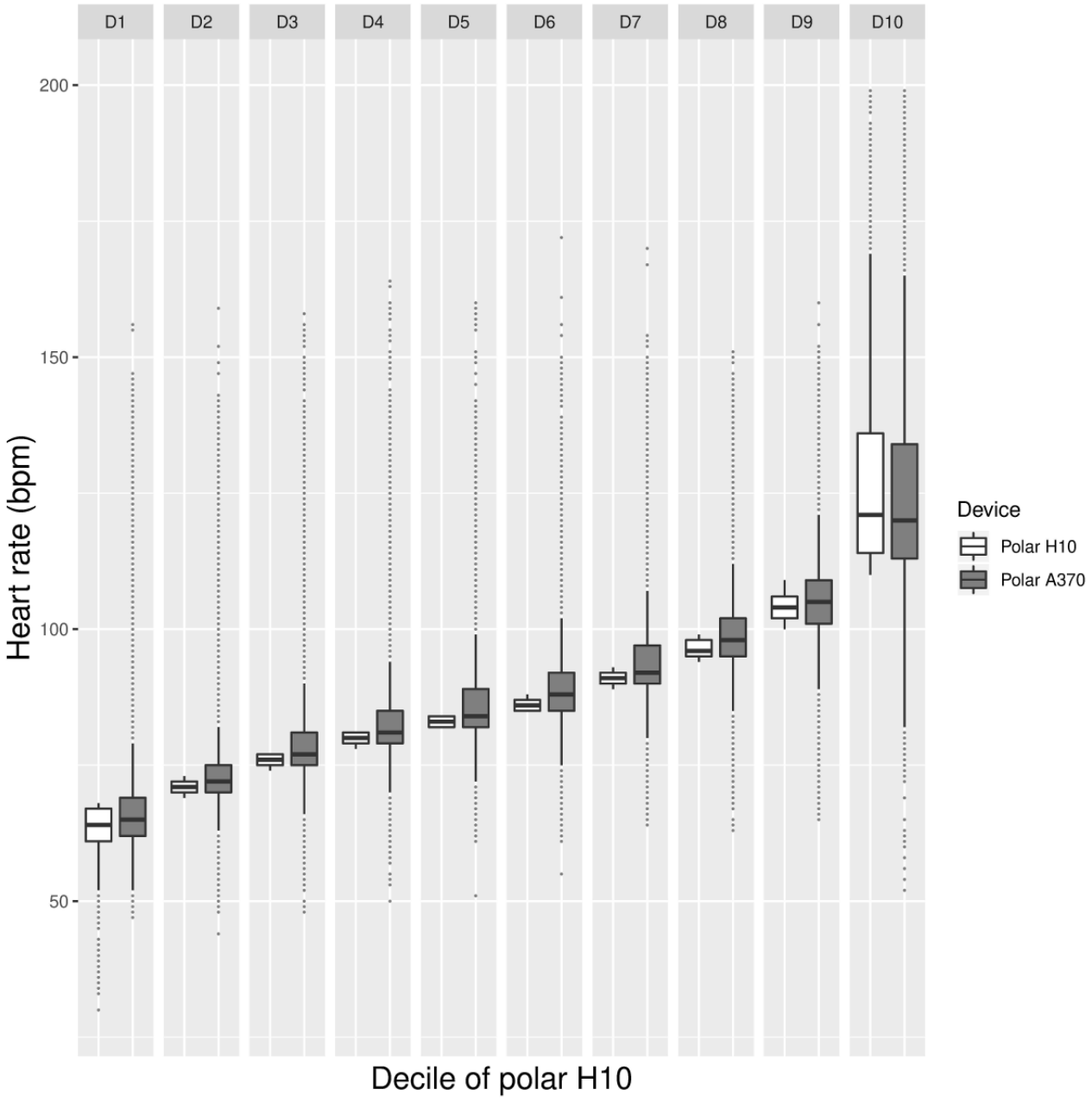


Figure 9. Free-living phase: box plot providing by-decile comparisons of mean Polar H10 (white) and Polar A370 HR data (gray). HR; heart rate; bpm: beats per minute.



Sensitivity and Specificity

When analyzing how many MVPA time points were identified by the Tempo HR and the Polar A370, we set the MVPA cutoff at 64% HR_{max}. In the laboratory phase, of the total aggregate time points in the MVPA HR zone that were detected by the Polar H10, 62.13% (1872/3013) were also identified by the Tempo HR, whereas the Polar A370 identified 81.09% (2273/2803). The remaining time was spent below the MVPA HR zone, of which 91.52% (4267/4662) and 97.52% (4637/4755) were also registered by the Tempo HR and the Polar A370, respectively. Overall, the Tempo HR identified 79.99% (6139/7675) and the Polar A370 91.42% (6910/7558) of data points accurately.

In the free-living phase, we found that the Tempo HR identified 54.27% (5717/10,535) and the Polar A370 identified 83.55% (9323/11,158) of the MVPA time points that the Polar H10 registered. The Tempo HR picked up 97.22% (186,402/191,741) and the Polar A370 picked up 96.72% (183,625/189,861) of time points below the MVPA HR zone. Overall accuracy was above 90% for both the trackers (Tempo HR: 94.98%, 192,119/202,276; Polar A370: 95.98%, 192,948/201,019). An overview of the results is provided in Table 1.

Table 1. Number of 10-second matched time points spent in heart rate zones as detected by the Polar A370 and the Tempo HR in the laboratory phase and free-living phase.

According to Polar H10	≥64% HR_{max}[a], n (%)	<64% HR_{max}, n (%)
Laboratory phase		
According to Polar A370		
≥64% HR_{max}	2273 (81.09)	118 (2.48)
<64% HR_{max}	530 (18.91)	4637 (97.52)
Total	2803 (37.09)	4755 (62.91)
According to Tempo HR		
≥64% HR_{max}	1872 (62.13)	395 (8.47)
<64% HR_{max}	1141 (37.87)	4267 (91.53)
Total	3013 (39.26)	4662 (60.74)
Free-living phase		
According to Polar A370		
≥64% HR_{max}	9323 (83.55)	6236 (3.28)
<64% HR_{max}	1835 (16.45)	183,625 (96.72)
Total	11,158 (5.55)	189,861 (94.45)
According to Tempo HR		
≥64% HR_{max}	5717 (54.27)	5339 (2.78)
<64% HR_{max}	4818 (45.73)	186,402 (97.22)
Total	10,535 (5.21)	191,741 (94.79)

[a]HR_{max}: maximum heart rate.

Discussion

Principal Findings

From the present 2-phased tracker validation study involving 55 participants with varying characteristics, a few key findings can be highlighted. First, HR data from the low-cost Tempo HR tracker showed moderate agreement with the data from the chest-strapped Polar H10 in both the laboratory phase and free-living phase. Although the measurement errors of the Tempo HR were above the 10% validity cutoff [46,47] in both phases, indicating its limited validity when measuring HR, the measurement error was markedly lower and close to the cutoff in the free-living phase (10.2% error). Second, HR data from the consumer-based Polar A370 showed strong agreement with data from the Polar H10 and low measurement errors (below the 10% validity cutoff) in both phases. Third, the differences between the Tempo HR and the Polar H10 are highest at higher HRs in both phases. This suggests that the measurement errors highlighted above are mainly the result of errors at high HRs. Further evidence for this conclusion can be derived from the sensitivity and specificity analysis where the Tempo HR identified only more than 50% of HRs above the MVPA threshold in both phases, whereas the Polar A370 identified more than 80% of HRs above the MVPA threshold in both phases. Fourth, agreement was generally higher and errors were smaller in the free-living phase compared with the laboratory phase. Finally, both trackers underestimated HR in the laboratory phase, whereas they overestimated it slightly in the free-living phase.

To establish the stability of the study results, we conducted sensitivity analyses. For this, we removed outliers and compared Polar H10 with the 2 other trackers using the remaining matched data points available. Outliers were defined as follows: a Pearson correlation coefficient of less than 0.3 between the Polar H10 and the test trackers in the laboratory setting. In secondary analyses, we only used data that were available from all 3 devices. Conducting these analyses did not change the results markedly (data not shown). As such, the reported results are not influenced by extreme cases or outliers.

Heart Rate Accuracy of the Polar A370 and Tempo HR in Context

When contextualizing our laboratory findings with those reported in the literature, the Polar A370 and the Tempo HR appear to have comparable or better accuracy with the market leader Fitbit, which has been studied extensively [20-25,28-32,34,36,37]. For example, authors who also asked participants to go through a cycling ergometer program reported agreement coefficients for Fitbit devices of between 0.21 and 0.50 [30,32,34]. The ICCs for the Polar A370 and Tempo HR in our study were 0.73 and 0.51, respectively. Similarly, other studies reported Fitbit MAPEs of 15.9% [34] and 21.06% [32], whereas the MAPE of the Polar A370 in our study was 6.4%; the one for the Tempo HR was 13.0%.

Comparing our results from the free-living phase with the results reported in other studies is problematic as, to our knowledge, there are only 2 studies that had a free-living element [38,39]. Gorny et al assessed data collected by the Fitbit Charge HR against data from the Polar H6 chest strap and reported an ICC of 0.83 and a mean difference between devices of −5.96 bpm. The ICC is in line with what we found for the Polar A370 in our study (ICC: 0.83). However, we observed overestimation in the free-living setting (Polar A370: 3.4 bpm; Tempo HR: 0.4 bpm). The authors also conducted sensitivity and specificity analysis and reported that the Fitbit Charge HR detected 52.9% of episodes spent in MVPA HR zones. Although this appears to be similar to what we found for the sensitivity of the Tempo HR (54.27%), the MVPA cutoffs in both studies were different. In our study, the more recent cutoff of 64% HR_{max} was used, whereas Gorny et al used the older 50% HR_{max} cutoff. One study by Nelson and Allen also provides some information on the accuracy of a Fitbit device in a free-living setting (Fitbit Charge 2). Over a 24-hour period, agreement measured by the concordance correlation coefficient was 0.91; this is close to what we found for the Polar A370 (although we used the ICC that provides similar estimates). MAE (4.9 bpm) and MAPEs (6.0%) for the Fitbit Charge 2 were also similar to that of the Polar A370 in our study (5.9 bpm, 7.1%). From these results, it appears that the Polar A370 is similarly accurate as the Fitbit Charge in free-living settings, whereas the Tempo HR appears to be less accurate.

The finding that the accuracy of wrist-worn trackers decreases as intensity increases has been observed in previous laboratory studies. For example, Boudreaux et al found that an increase in cycling intensity was associated with increasing HR underestimation in assessed activity trackers [32]. Dondzila et al made a similar discovery during treadmill exercises [21]. Spierer et al suggested that the increased measurement error with increasing movement intensity is because of increased motion, which leads to more disturbances of the blood flow-sensor interface [35].

It is difficult to draw firm conclusions about such trends in the free-living phase, as there are no comparable studies available. We observed smaller differences across activity intensities, which might be related to the fact that the proportion of higher HR values was rather small compared with the laboratory study. This might also partially explain the generally higher accuracy in the free-living phase versus the laboratory phase. Another reason for the difference in accuracy between the free-living phase and laboratory phase might be related to the temperature difference between the laboratory and the free-living settings [51]. The laboratory study was conducted in an air-conditioned environment in which the temperature varied between 18°C and 20°C. This is significantly colder than the outside temperature in Singapore (between 30°C and 32°C); hence, the free-living study was executed under warmer conditions. The fact that higher temperatures facilitate blood flow is well established. The aforementioned factors could also partially explain why the test devices underestimated HR in the laboratory phase and not in the free-living phase.

Differences in Accuracy Between Devices

From the results of our study and the overall HR tracker validation literature, it is obvious that there are marked differences between devices in terms of accuracy that ought to be explained. A review by Tamura et al provides some insights into the factors that impact HR measurement through PPG in different devices [19]. First, PPG-measured HR differences between devices might be related to the algorithms used to estimate HR. Different devices use different algorithms for translating the detected blood flow into HR. A recent study highlighted that sensor technologies to detect physiological parameters are mostly identical between devices. However, the algorithms applied to translate the collected data into a readable HR measure vary from vendor to vendor and can be changed without notice [14]. Similarly, algorithms used to correct for movement artifacts during upper body movements vary between devices [52]. As such, it is the algorithm and not the technology itself that seems to primarily impact device accuracy. A second reason for the observed differences in accuracy between devices could be related to the contact force between the sensor and the skin [53]. Insufficient contact pressure is related to less sensitivity in detecting blood flow. Although both trackers were fitted snuggly (this was tested), the Polar A370 had more bracelet holes, which meant that its sensor might have had slightly better contact with the skin.

Strengths and Limitations

A number of strengths of this study can be highlighted. To the best of our knowledge, this is the first study that thoroughly investigated the validity of HR measures of modern wrist-worn activity trackers in 2 settings, the laboratory and daily life. Research on the real-world performance of activity trackers can advance the PA and exercise measurement field substantially as these trackers are meant to be used as people go about their normal lives. Second, our study sample size was relatively large and diverse, which is rare in validation studies. Third, we were able to collect temporally dense HR data from all devices (approximately 12 hours per device in the free-living phase), which allowed us to conduct in-depth analyses of tracker validity across varying HRs. The richness of data we collected stands in stark contrast to most previous studies that relied mainly on few data points, for example, at the end or midpoint of a stage in a cycling protocol [32]. Despite these strengths, a few limitations ought to be mentioned. First, we opted for a cycling protocol in our laboratory phase, which might not be optimal as participants bent their wrists when holding on to the handlebar. However, using other protocols, such as a treadmill program, as in other studies, is not optimal either, as upper body movements will lead to movement artifacts that are likely to impact HR measures of the wrist-worn trackers [19]. Second, compared with other researchers who investigated the validity of up to 8 activity trackers [32], we only used 2 trackers in our study. Although this might appear to be a significant shortcoming, we limited the number of trackers intentionally. We based our decision on a small internal pilot study during which we established that wearing many trackers in addition to a chest-strap HR monitor during a free-living study would be too burdensome for participants; as such, we were concerned about study compliance. Third, it was not possible to ensure

participants wore the devices accurately during the free-living phase. Although we explained how the devices should be fitted, practiced the wear protocol with participants, and provided a step-by-step instruction sheet, it is possible that participants did not wear the devices appropriately. However, as the accuracy was generally higher in the free-living phase versus the laboratory phase, we believe that inappropriate wear did not introduce significant errors. Fourth, we predetermined the wearing side of the 2 trackers (Tempo HR: left and Polar A370: right). There is some debate about whether the side at which trackers are worn could influence the HR data collected. Some researchers predetermined wear side [21,30,31,36-38,40], whereas others used randomization procedures [20,22,24,29,32]. We believe our protocol did not introduce bias and base this assumption on a 2017 study in which researchers found that the wearing side (left or right) was not associated with differences in HR measurement error in 6 commercial trackers; in 1 device, there were some small differences [28]. Finally, we used a chest-strapped HR monitor as our criterion device for measuring HR. Although ECG might have been the more adequate criterion, Polar chest straps are generally accepted reference devices as they have adequate validity [42]. In addition, Polar chest straps were used in many previous studies, and they were the only feasible criterion in our free-living phase.

Future Directions

A recent review highlighted the strong increase in the availability and use of wrist-worn activity trackers and identified 432 different activity trackers that belonged to 123 unique brands [14]. As such, researchers will increasingly make use of them [11,54]. A key promise of such wrist-worn trackers is that they can facilitate PA behavior change through self-monitoring and feedback, 2 well-established behavioral change techniques that are supposed to enable individuals to bridge the gap between current behavior and behavioral targets [55,56]. However, research evidence on the effectiveness of tracker-based self-monitoring and feedback in terms of PA is currently mixed [13,57]. The effect of tracker use on PA behavior might be moderated by actual wear time [58]. In addition to self-monitoring, activity trackers are proposed to be important for just-in-time adaptive interventions in which in-the-moment behavioral support is delivered based on real-life data. Step counts are most commonly used for this purpose, with HR data as a basis for just-in-time support and feedback being a viable option. Due to the ability of sensors to communicate sensor-collected information on PA intensity to mobile phone apps, real-time adaptations of feedback and support is possible. Such dynamic interventions are suggested to increase sustainable behavior change through effective engagement [54,59]. Research in this field is in its infancy, but important gains are being made. Finally, observational research is likely to be a great beneficiary of tracker devices as they can be used to collect long-term PA data in an unobtrusive and resource-effective way.

Acknowledgments

This research is supported by the Singapore Ministry of Health's National Medical Research Council under the Fellowship Programme by Singapore Population Health Improvement Centre (NMRC/CG/C026/2017_NUHS). The authors would like to acknowledge all participants who took part in the study. Finally, AMM acknowledges his new-born daughter, Lia Yihan, who was so kind to only cry after a few reviewer comments were addressed.

Authors' Contributions

AMM, NXW, and FMR conceived the study. Data collection was conducted by AMM and NXW. ICCL provided expertise for the laboratory study and supported the setup of the study. JY and CST analyzed the data with iterative feedback from AMM, NXW, and FMR. NL, JT, and AT supported data extraction and provided critical feedback throughout. AMM wrote the manuscript and received feedback from all coauthors. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of raw data retrieved from the Polar H10, Polar A370, and Tempo HR.
[XLSX File (Microsoft Excel File)11 KB - mhealth_v7i10e14120_app1.xlsx]

References

1. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Lancet Physical Activity Series 2 Executive Committee. Progress in physical activity over the Olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
2. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. Lancet 2016 Sep 24;388(10051):1337-1348 [FREE Full text] [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]

3. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, et al. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. Lancet 2017 Dec 16;390(10113):2643-2654. [doi: 10.1016/S0140-6736(17)31634-3] [Medline: 28943267]
4. Lee IM, Shiroma EJ, Evenson KR, Kamada M, LaCroix AZ, Buring JE. Accelerometer-measured physical activity and sedentary behavior in relation to all-cause mortality: the women's health study. Circulation 2018 Jan 9;137(2):203-205 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.031300] [Medline: 29109088]
5. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. Br Med J 2016 Aug 9;354:i3857 [FREE Full text] [doi: 10.1136/bmj.i3857] [Medline: 27510511]
6. Kim J, Im JS, Choi YH. Objectively measured sedentary behavior and moderate-to-vigorous physical activity on the health-related quality of life in US adults: the National Health and Nutrition Examination Survey 2003-2006. Qual Life Res 2017 May;26(5):1315-1326. [doi: 10.1007/s11136-016-1451-y] [Medline: 27837382]
7. White RL, Babic MJ, Parker PD, Lubans DR, Astell-Burt T, Lonsdale C. Domain-specific physical activity and mental health: a meta-analysis. Am J Prev Med 2017 May;52(5):653-666. [doi: 10.1016/j.amepre.2016.12.008] [Medline: 28153647]
8. Skender S, Ose J, Chang-Claude J, Paskow M, Brühmann B, Siegel EM, et al. Accelerometry and physical activity questionnaires - a systematic review. BMC Public Health 2016 Jun 16;16:515 [FREE Full text] [doi: 10.1186/s12889-016-3172-0] [Medline: 27306667]
9. Butte NF, Ekelund U, Westerterp KR. Assessing physical activity using wearable monitors: measures of physical activity. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S5-12. [doi: 10.1249/MSS.0b013e3182399c0e] [Medline: 22157774]
10. 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]
11. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related ehealth and mhealth research: bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
12. Wilde LJ, Ward G, Sewell L, Müller AM, Wark PA. Apps and wearables for monitoring physical activity and sedentary behaviour: a qualitative systematic review protocol on barriers and facilitators. Digit Health 2018;4:2055207618776454 [FREE Full text] [doi: 10.1177/2055207618776454] [Medline: 29942637]
13. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: a systematic review. BMC Public Health 2015 Jun 24;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
14. Henriksen A, Mikalsen MH, 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]
15. Sartor F, Papini G, Cox LG, Cleland J. Methodological shortcomings of wrist-worn heart rate monitors validations. J Med Internet Res 2018 Jul 2;20(7):e10108 [FREE Full text] [doi: 10.2196/10108] [Medline: 29967000]
16. 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]
17. 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]
18. 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]
19. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable photoplethysmographic sensors—past and present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]
20. 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]
21. Dondzila CJ, Lewis C, Lopez JR, Parker T. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. Int J Exerc Sci 2018;11(7):575-584 [FREE Full text]
22. 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]
23. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. 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]

24. 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;4:2055207618770322 [FREE Full text] [doi: 10.1177/2055207618770322] [Medline: 29942628]
25. 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 1;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
26. Delgado-Gonzalo R, Parak J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:430-433. [doi: 10.1109/EMBC.2015.7318391] [Medline: 26736291]
27. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]
28. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):pii: E3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
29. 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]
30. 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]
31. Bai Y, Hibbing P, Mantis C, Welk GJ. 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]
32. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, 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]
33. Parak J, Uuskoski M, Machek J, Korhonen I. Estimating heart rate, energy expenditure, and physical performance with a wrist photoplethysmographic device during running. JMIR Mhealth Uhealth 2017 Jul 25;5(7):e97 [FREE Full text] [doi: 10.2196/mhealth.7437] [Medline: 28743682]
34. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, 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]
35. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
36. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep;15(3):540-547 [FREE Full text] [Medline: 27803634]
37. 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]
38. 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]
39. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]
40. Sartor F, Gelissen J, van Dinther R, Roovers D, Papini GB, Coppola G. Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients. BMC Sports Sci Med Rehabil 2018;10:10 [FREE Full text] [doi: 10.1186/s13102-018-0098-0] [Medline: 29881626]
41. Shephard RJ. Qualified fitness and exercise as professionals and exercise prescription: evolution of the PAR-Q and Canadian aerobic fitness test. J Phys Act Health 2015 Apr;12(4):454-461. [doi: 10.1123/jpah.2013-0473] [Medline: 24836975]
42. Cheatham SW, Kolber MJ, Ernst MP. Concurrent validity of resting pulse-rate measurements: a comparison of 2 smartphone applications, the polar H7 belt monitor, and a pulse oximeter with Bluetooth. J Sport Rehabil 2015;24(2):171-178. [doi: 10.1123/jsr.2013-0145] [Medline: 24700420]
43. Fox SM, Naughton JP. Physical activity and the prevention of coronary heart disease. Prev Med 1972 Mar;1(1-2):92-120. [doi: 10.1016/0091-7435(72)90079-5]
44. Borg G. Borg's Perceived Exertion and Pain Scales. Champaign, Illinois: Human Kinetics; 1998.
45. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
46. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. 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]
47. DocPlayer. 2002. AAMI. American National Standard. Cardiac monitors, heart rate meters, and alarms ANSI/AAMI EC13:2002 URL: https://docplayer.net/

34982183-Aami-american-national-standard-cardiac-monitors-heart-rate-meters-and-alarms-ansi-aami-ec13-2002.html [accessed 2019-08-08]
48. Howley ET. Type of activity: resistance, aerobic and leisure versus occupational physical activity. Med Sci Sports Exerc 2001 Jun;33(6 Suppl):S364-9; discussion S419. [doi: 10.1097/00005768-200106001-00005] [Medline: 11427761]
49. US Department of Health & Human Services, Centers for Disease Control and Prevention. Physical Activity And Health: A Report Of The Surgeon General. Atlanta, Georgia: US Department of Health & Human Services; 1996.
50. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. Tenth Edition. Baltimore, Maryland: Lippincott Williams and Wilkins; 2017.
51. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
52. Zhou C, Feng J, Hu J, Ye X. Study of artifact-resistive technology based on a novel dual photoplethysmography method for wearable pulse rate monitors. J Med Syst 2016 Mar;40(3):56. [doi: 10.1007/s10916-015-0412-2] [Medline: 26645320]
53. Teng XF, Zhang YT. The effect of contacting force on photoplethysmographic signals. Physiol Meas 2004 Aug 12;25(5):1323-1335. [doi: 10.1088/0967-3334/25/5/020]
54. 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]
55. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 4;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
56. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
57. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: a systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
58. Hartman SJ, Nelson SH, Weiner LS. Patterns of Fitbit use and activity levels throughout a physical activity intervention: exploratory analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2018 Feb 5;6(2):e29 [FREE Full text] [doi: 10.2196/mhealth.8503] [Medline: 29402761]
59. Hekler EB, Rivera DE, Martin CA, Phatak SS, Freigoun MT, Korinek E, et al. Tutorial for using control systems engineering to optimize adaptive mobile health interventions. J Med Internet Res 2018 Jun 28;20(6):e214 [FREE Full text] [doi: 10.2196/jmir.8622] [Medline: 29954725]

Abbreviations

BA: Bland-Altman
BMI: body mass index
bpm: beats per minute
ECG: echocardiogram
HPB: Health Promotion Board
HR: heart rate
HR_{max}: maximum heart rate
ICC: intraclass correlation coefficient
LoA: limits of agreement
MAE: mean absolute error
MAPE: mean absolute percentage error
MVPA: moderate-to-vigorous physical activity
NSC: National Steps Challenge
PA: physical activity
PPG: photoplethysmography

Edited by G Eysenbach; submitted 24.03.19; peer-reviewed by F Sartor, G Signorelli, X Ye; comments to author 29.04.19; revised version received 21.06.19; accepted 08.07.19; published 02.10.19.

Please cite as:
Müller AM, Wang NX, Yao J, Tan CS, Low ICC, Lim N, Tan J, Tan A, Müller-Riemenschneider F
Heart Rate Measures From Wrist-Worn Activity Trackers in a Laboratory and Free-Living Setting: Validation Study
JMIR Mhealth Uhealth 2019;7(10):e14120
URL: https://mhealth.jmir.org/2019/10/e14120
doi:10.2196/14120
PMID:31579026

Original Paper

Accuracy of the Multisensory Wristwatch Polar Vantage's Estimation of Energy Expenditure in Various Activities: Instrument Validation Study

Rahel Gilgen-Ammann[1], PhD; Theresa Schweizer[1], BSc; Thomas Wyss[1], PhD

Swiss Federal Institute of Sport Magglingen, Magglingen, Switzerland

Corresponding Author:
Rahel Gilgen-Ammann, PhD
Swiss Federal Institute of Sport Magglingen
Hauptstrasse 247
Magglingen, 2532
Switzerland
Phone: 41 584676321
Email: rahel.gilgen@baspo.admin.ch

Abstract

Background: Sport watches and fitness trackers provide a feasible way of obtaining energy expenditure (EE) estimations in daily life as well as during exercise. However, today's popular wrist-worn technologies show only poor-to-moderate EE accuracy. Recently, the invention of optical heart rate measurement and the further development of accelerometers in wrist units have opened up the possibility of measuring EE.

Objective: This study aimed to validate the new multisensory wristwatch Polar Vantage and its EE estimation in healthy individuals during low-to-high-intensity activities against indirect calorimetry.

Methods: Overall, 30 volunteers (15 females; mean age 29.5 [SD 5.1] years; mean height 1.7 [SD 0.8] m; mean weight 67.5 [SD 8.7] kg; mean maximal oxygen uptake 53.4 [SD 6.8] mL/min·kg) performed 7 activities—ranging in intensity from sitting to playing floorball—in a semistructured indoor environment for 10 min each, with 2-min breaks in between. These activities were performed while wearing the Polar Vantage M wristwatch and the MetaMax 3B spirometer.

Results: After EE estimation, a mean (SD) of 69.1 (42.7) kcal and 71.4 (37.8) kcal per 10-min activity were reported for the MetaMax 3B and the Polar Vantage, respectively, with a strong correlation of r=0.892 (P<.001). The systematic bias was 2.3 kcal (3.3%), with 37.8 kcal limits of agreement. The lowest mean absolute percentage errors were reported during the sitting and reading activities (9.1%), and the highest error rates during household chores (31.4%). On average, 59.5% of the mean EE values obtained by the Polar Vantage were within ±20% of accuracy when compared with the MetaMax 3B. The activity intensity quantified by perceived exertion (odds ratio [OR] 2.028; P<.001) and wrist circumference (OR −1.533; P=.03) predicted 29% of the error rates within the Polar Vantage.

Conclusions: The Polar Vantage has a statistically moderate-to-good accuracy in EE estimation that is activity dependent. During sitting and reading activities, the EE estimation is very good, whereas during nonsteady activities that require wrist and arm movement, the EE accuracy is only moderate. However, compared with other available wrist-worn EE monitors, the Polar Vantage can be recommended, as it performs among the best.

(JMIR Mhealth Uhealth 2019;7(10):e14534) doi:10.2196/14534

KEYWORDS

validation; mHealth and eHealth; activity monitor

Introduction

Previous Research

The accurate measurement of human body energy expenditure (EE) is an important parameter for many applications [1,2]. For example, the intensity of physical activity can be evaluated based on the energy consumed during exercise, and dietary guidance can be given based on the total daily EE [3].

The exact measurement of human EE requires laboratory methods that are not suitable for performing in everyday life,

such as wearing a face mask that measures respiration gases or analyzing saliva or urine samples using expensive doubly labeled water techniques [4]. Body-worn sensors provide a consumer-friendly option for measuring EE in daily life and during exercise; originally, these sensors used heart rate (HR) to estimate EE during exercise [5,6] or were assisted by accelerometer-based measurements. Recently, however, the invention of optical HR measurement and the further development of accelerometers in wrist units have opened up the new possibility of feasibly measuring EE in exercise and in daily life [4,7]. These technologies rely on photoplethysmography and use HR-derived algorithms to contribute to the estimation of EE based on activity intensity [8,9]. However, previous validation studies on a variety of body-worn sensor brands reported error rates of approximately 10% to 210%, with more accurate values measured during high-intensity aerobic exercises, such as running or cycling, than during daily activities of low-to-moderate intensity, such as lying down or sitting [10-13]. The limitations of such sensors include distorted optical pulse signals because of motion artifacts, the inability of the devices to account for additional load carried by the user, and distal sensor placement on the body [7,14].

This Study

Smart Calories is a novel EE calculation by Polar that aims to improve the EE estimation in daily life and exercise. The Polar Vantage (V and M) is a multisensory wrist-worn technology that expresses EE in calories per activity. This new measurement system has not been validated before. Therefore, the aim of this study was to validate Polar Vantage and its EE estimation in a healthy and heterogeneous sample at rest and during different exercise modes and at different intensities against the criterion of indirect calorimetry measure.

Methods

Participants

A total of 30 healthy and lean volunteers gave informed consent to participate in this study. None of the participants were known to be taking any medications affecting HR or metabolism nor did they have any tattoos on the nondominant wrist. Of the participants, 50% (15/30) were female and all were within the age range of 20 to 40 years. In terms of activity levels, 33% (10/30) of the participants met or nearly met the physical activity guidelines (ie, completing 150 min of moderate-intensity activity per week), 33% (10/30) of the participants were active (ie, participating in regular training but with no competitive targets), and 33% (10/30) of the participants were endurance athletes (ie, regularly participating in running, triathlon, or cycling competitions). The participants completed written informed consent forms and physical activity readiness questionnaires (PAR-Q) before taking part in the study. None answered *yes* to any PAR-Q question. For sample size estimation, the data of a similar study and similar expected monitor accuracy were used as the estimate for a paired 2-tailed *t* test [13]. Approval for this study was granted by the ethics commission of the Canton Berne (2018-00309), and it conformed to the principles of the Declaration of Helsinki.

Procedure

The measurements were taken on 2 separate test days. On day 1, the resting HR, maximal oxygen uptake (VO_{2max}), and maximal heart rate (HR_{max}) were obtained in a laboratory. The participants were scheduled during the morning hours, and they were instructed to avoid any strenuous physical activity and caffeine intake for a minimum of 24 and 12 hours, respectively, before the appointment. First, information about the study was verbally repeated to the participants, and both the questionnaires and informed consent form were confirmed. Second, the body weight (of the participant in underwear), body height, wrist circumference, skin color (using the Fitzpatrick scale from 1 lightest tone to 6 darkest) [15], and skin hair on the wrist (0=little hair, 1=moderate or a lot of hair) were assessed by the supervisor. Third, an HR strap was mounted around the participant's chest. Then, the HR was measured for 10 min with the participant in a supine position and completely at rest in a quiet and thermoneutral environment (20°-22° C) [16]. Thereafter, the VO_{2max} and HR_{max} assessments were conducted on a treadmill (h/p/cosmos pulsar; Cosmos Sports & Medical Ltd). Initially, a warm-up and treadmill familiarization period of 5 min at 8.6 km/h and 0% inclination was conducted, followed by a short rest period during which a spirometer face mask was fixed on the participant. To determine the VO_{2max} and HR_{max}, a graded protocol from an initial speed of 7.5 km/h and a 7% constant inclination was applied, with a speed increase of 0.5 km/h every 30 seconds until voluntary exhaustion [17]. Immediately after voluntary exhaustion, the participant was asked to rate the perceived exertion using the Borg scale (6-20) [18]. For the determination of VO_{2max}, at least 2 of the following 4 maximum criteria had to be fulfilled: respiratory exchange ratio greater than 1.1, voluntary maximum (Borg scale ≥ 18), plateau in VO_2, or HR greater than 85% HR_{max} (HR_{max} estimation based on 220 bpm minus the participant's age in years) [17].

On day 2, the measurements of EE values during rest, daily activities, and sport activities were obtained in a gym hall with prepared areas. The participants were scheduled anytime during the day, 2 hours after the last food intake, and 12 hours after the last caffeine intake. The participants were equipped with the Polar Vantage M, the Polar H10 chest strap, the MetaMax 3B data logger, and a spirometer face mask. Following a short recreational walk of 3 min as a warm-up, the participants performed 7 simulated free-living activities for 10 min each. After each activity, the participant walked to the next activity and stood still for the remainder of the 2-min break; however, following the running activity, the break lasted for 4 min. During these recovery phases, the upcoming activity was explained. The starting and stopping time of each activity task was registered by the researcher on a paper version of the study protocol, using a master stopwatch. Moreover, when each activity was started and stopped, the respective EE values given by the Polar Vantage M for each activity were recorded.

The 7 activities and their order were as follows:

1. Sitting in a chair and reading (sedentary activity; training mode *other indoor*).

2. Wiping the floor with a mop and hanging out the laundry at a self-guided order and pace (household chores; training mode *other indoor*).
3. Normal walking on an indoor round track of 290 meters with the pacing instruction "as you would go to the bus station in no rush" (gait activity; training mode *walking*).
4. Jogging on an indoor round track of 290 meters with the pacing instruction "choose your own pace at which you could talk to someone" (gait activity; training mode *running*).
5. A strength training circuit of 45-second workouts with a dumbbell in each hand followed by 15-second rests, including squats, shoulder shrugs, bicep curls, lunges, and sit-ups [19] with the instruction "choose your own dumbbell weight and pacing so that the workout is at least somewhat hard" (Borg scale value >12; sport activity; training mode *strength training*).
6. Cycling on an ergometer (Ergoselect 200; Ergoline GmbH) at 80 rounds per min and an HR around 120 beats per min (sport activity; training mode *indoor cycling*).
7. A floorball course (approximately 80 meters in length) including drippling, passing the ball, shooting, and jogging [20], for which the task execution was self-paced but short recovery phases of 10 seconds slow walking per round were required (sport activity; training mode *other indoor*).

Immediately after the termination of each task, individual Borg scale values were reported to rate the perceived exertion [18].

Instruments

The investigated device was the Polar Vantage M wristwatch (Polar Electro Oy), which uses a bioimpedance-assisted optical HR calculation and 3D acceleration signal. The Polar Vantage M was placed on the participant's nondominant wrist, 1 finger behind the wrist bone. The participant's anthropometrics, resting HR, HR_{max}, and VO_{2max} values were entered into the user profile, and each corresponding training mode was set in the user setting before starting the respective activity. HR was assessed using the Polar H10 chest strap [21]. To obtain measures of oxygen consumption (VO_2) and carbon dioxide production (VCO_2) to determine the VO_{2max}, the Quark CPET (Cosmed) was used. To calculate the EE criterion during simulated free-living activities, the VO_2 and VCO_2 were obtained using a portable open-circuit metabolic system (MetaMax 3B; Cortex Biophysik) [22,23]. The MetaMax 3B was mounted on the participant with a face mask and a chest harness. All devices were calibrated before each measurement according to the manufacturer's instructions.

Data Processing

Each participant's resting HR was calculated based on the average minimum 30-second values obtained during the 10-min resting measurement, whereas the VO_{2max} and HR_{max} were calculated based on the average maximal 30-second values obtained during the graded treadmill test [17].

To investigate the EE estimations on measurement day 2, the EE values shown on the Polar Vantage M display were noted for each single 10-min activity. To calculate the EE criterion, the formula presented by Elia and Livesey [24] was used to sum up the gas exchange data in kilocalories per minute to generate the total EE per 10-min activity [25]. Each of the 30 participants completed all 7 activity tasks, but 2 technical failures during the floorball course were reported. Therefore, of the 210 activities, 208 were recorded and analyzed (99%).

Statistical Analysis

Descriptive statistics with mean absolute and percentage errors, Pearson correlations, Bland and Altman analysis, and EE 20% accuracy were used. Bland and Altman analyses with corresponding 95% limits of agreement (SD 1.96) were used to calculate and visualize systematic differences in the EE estimations [26]. The EE 20% accuracy was defined as the percentage at which the Polar Vantage M was within the proposed equivalence zone of ± 20% from the criterion values [12,19,27]. Moreover, backward multiple linear regression analyses with the mean absolute error (MAE) as the dependent variable were performed to investigate potential confounding effects of the independent variables of gender, body mass index, wrist circumference, skin color, wrist hair, HR, resting HR, HR_{max}, VO_{2max}, and perceived exertion (Borg scale value) on EE accuracy. In addition, backward multiple linear regression analyses were performed on clustered activity groups: (1) low-to-moderate-intensity activities (sitting and reading, household chores, and walking) and high-intensity activities (jogging, strength training circuit, cycling, and floorball course) and (2) activities with no-to-little (steady) arm movement (sitting and reading, walking, jogging, and cycling) and activities with a lot (unsteady) of arm movement (household chores, strength training circuit, and floorball course). In the case of multicollinearity ($r \geq 0.80$) or the nonsignificant prediction of the MAE, the relevant variable was excluded from the regression analysis. Any P value less than .05 was considered statistically significant.

Results

Each participant's characteristics are presented in Table 1. The MetaMax 3B and Polar Vantage M reported the mean (SD) of EE to be 69.1 (42.7) kcal and 71.4 (37.8) kcal per 10-min activity, respectively (Table 2), with a correlation of r=0.892 (P<.001). Measured EE ranged from 10 to 194 kcal per 10 min, with the highest EE values obtained during the floorball course activity and the lowest EE values obtained during sitting and reading. The systematic bias was 2.3 kcal (3.3%), with 37.8 kcal limits of agreement (Figure 1). The mean absolute percentage error (MAPE) of the Polar Vantage M was 20.6%, ranging from 9.1% to 31.4%. On average, 59.5% of the mean EE values were accurate to within 20% when compared with those of the MetaMax 3B (Table 2). The household chores revealed the lowest accuracy (26.7%), whereas the sitting and reading revealed the highest accuracy (93.3%; Figure 2). Owing to its multicollinearity with perceived exertion (r=0.866; P<.001), the variable HR had to be excluded. A significant regression equation was revealed ($F_{2,207}$=42.628; P<.001), with an R^2 of 0.29. The perceived exertion (Borg scale value; odds ratio [OR] 2.028; P<.001) and wrist circumference (OR −1.533; P=.03) predicted 29% of the MAE within the Polar Vantage M.

Both predictors induced an underestimation of EE. The final linear regression models for clustered activity groups were similar to the data presented for the overall activities. The remaining independent variables, perceived exertion and wrist circumference, explained about 20% and 15% of the MAE in clustered low-to-moderate-intensity and high-intensity activities as well as 44% and 17% of the MAE in activities with no-to-little (steady) arm movement and a lot (unsteady) of arm movement.

Table 1. Participants' characteristics.

Characteristics	Female (n=15), mean (SD)	Male (n=15), mean (SD)	Differences	*P* value	Overall (N=30), mean (SD)
Age (years)	30.1 (1.7)	28.9 (1.8)	−1.3	.51	29.5 (5.1)
Body height (m)	1.7 (0.1)	1.8 (0.1)	0.1	<.001	1.7 (0.8)
Body weight (kg)	60.2 (4.5)	74.8 (4.7)	14.7	<.001	67.5 (8.7)
Body mass index (kg/m^2)	21.1 (1.3)	23.5 (1.5)	2.4	<.001	22.3 (1.8)
Wrist circumference (cm)	15.5 (0.8)	17.1 (0.9)	1.6	<.001	16.3 (1.2)
Skin hair on the wrist[a]	0.0 (0.0)	0.3 (0.5)	−0.3	<.001	0.2 (0.4)
Skin color[b]	2.9 (0.4)	2.9 (0.4)	0.0	>.99	2.9 (0.4)
Maximal oxygen uptake (mL/min·kg)	49.7 (6.0)	57.2 (5.5)	7.5	<.001	53.4 (6.8)
Resting heart rate (bpm[c])	54 (10)	53 (7)	−0.7	.82	54 (9)
Maximal heart rate (bpm)	181 (11)	189 (9)	7.8	.04	185 (11)

[a]0=little hair, 1=moderate or a lot of hair.

[b]Fitzpatrick scale 1 to 6.

[c]bpm: beats per minute.

Table 2. Perceived exertion, heart rate, measured energy expenditure, and error rates of the Polar Vantage M when compared with the MetaMax 3B.

Activity	Borg scale (6-20), mean (SD)	% HR_{max}[a]	EE[b] by MetaMax 3B (kcal), mean (SD)	EE by Polar Vantage M (kcal), mean (SD)	Systematic bias in kcal, (limits of agreement)	Mean absolute error in kcal, (mean absolute percentage error)	20% accuracy[c], n/N (%)
All activities	10.9 (3.5)	59.6	69.1 (42.7)	71.4 (37.8)	2.3 (37.8)	14.0 (20.6)	123/208 (59.1)
Sitting and reading	6.1 (0.3)	34.0	13.7 (2.5)	13.6 (2.1)	−0.1 (3.1)	1.2 (9.1)	28/30 (93)
Household chores	7.6 (1.0)	45.7	39.1 (8.6)	50.1 (10.6)	11.0 (16.5)	11.8 (31.4)	8/30 (27)
Walking (5.4 [0.5] km/h)	8.5 (0.9)	46.3	43.4 (6.8)	52.3 (8.5)	9.0 (12.3)	9.0 (21.4)	15/30 (50)
Jogging (9.6 [1.2] km/h)	11.8 (1.0)	71.4	113.7 (26.7)	103.3 (21.5)	−10.4 (33.5)	16.9 (14.8)	21/30 (70)
Strength training circuit (5.1 [2.8] kg)	13.5 (0.9)	63.3	61.8 (16.9)	71.0 (18.5)	9.2 (29.7)	14.3 (25.2)	15/30 (50)
Cycling (130.7 [22.6] W)	14.1 (1.1)	72.9	90.2 (15.5)	102.0 (24.9)	11.8 (34.9)	18.0 (20.6)	20/30 (66.7)
Floorball course	15.1 (1.7)	83.4	125.5 (36.1)	110.1 (29.3)	−15.4 (63.0)	27.5 (21.8)	16/28 (57.1)

[a]%HR_{max}=percentage of maximal heart rate.

[b]EE: energy expenditure.

[c]Percentage at which the EE estimated by the Polar Vantage M was within 20% from the criterion MetaMax 3B.

Figure 1. Bland and Altman plot of the energy expenditure (EE) estimation obtained during the 7 activity tasks (208 measurements). The solid line represents the systematic bias; the dashed lines represent the limits of agreement (systematic bias, SD 1.96).

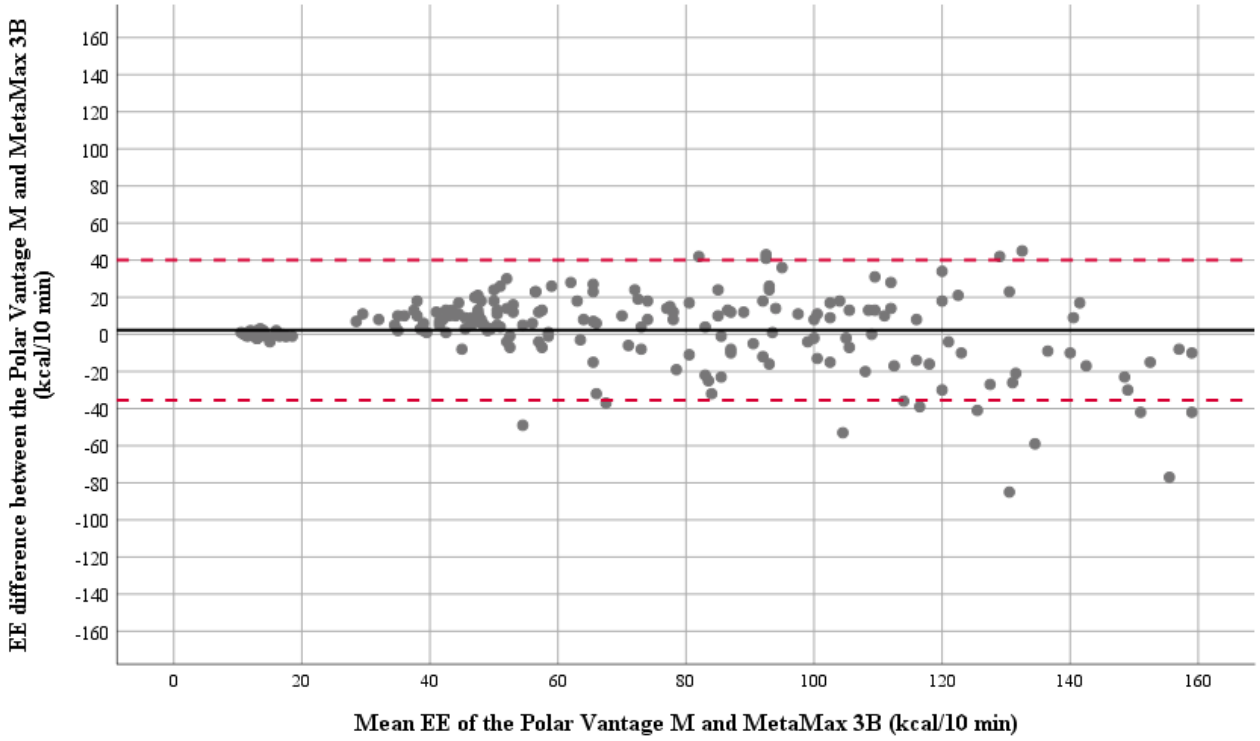
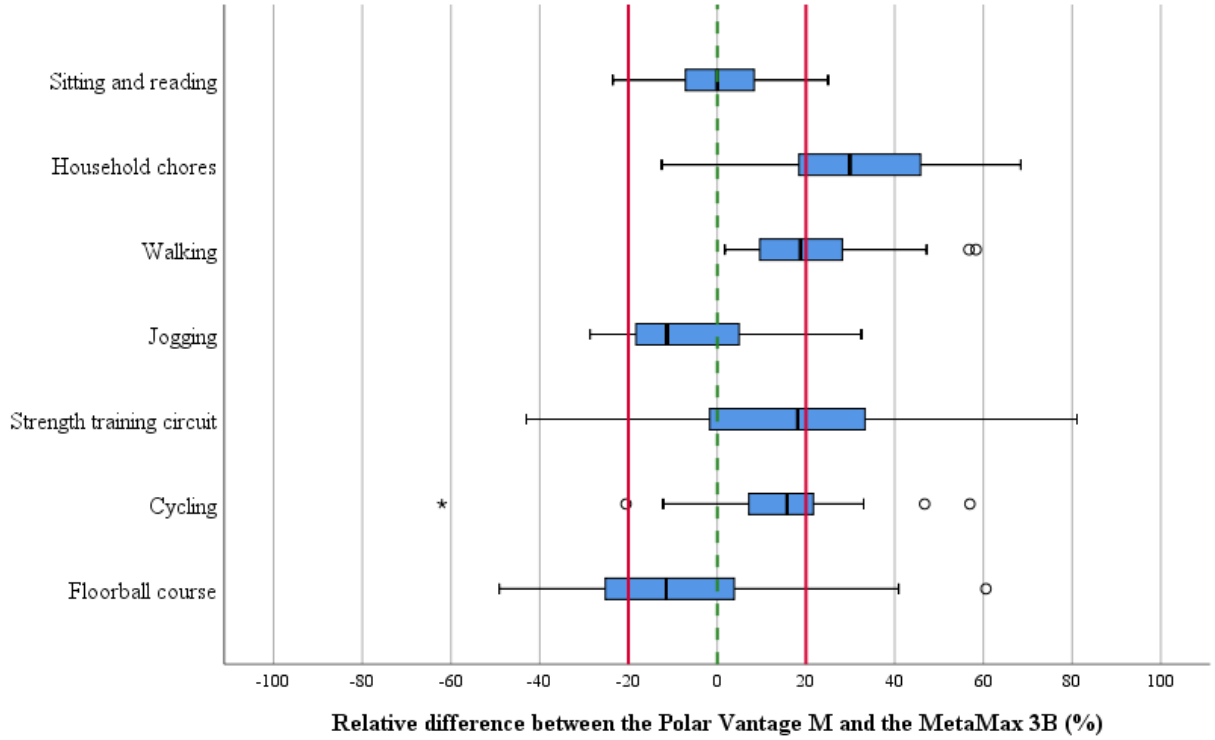


Figure 2. Relative deviation of the energy expenditure (EE) estimation by the Polar Vantage M compared with the criterion measurement of the MetaMax 3B separated for each activity task. The red lines indicate the proposed equivalence zone (SD 20% of the mean); the lower and upper boundary of the boxplots indicate the 25% and 75% quantiles of EE data, respectively, and the middle notch indicates the median data value. The whiskers include all data points that fall within the 1.5 interquartile range of the 25% and 75% quantile values. Circles and stars indicate EE data points that lie beyond the 1.5 and 3 interquartile ranges, respectively.



Discussion

Principal Findings

In this study, a recently launched multisensory wristwatch, the Polar Vantage, was evaluated. The accuracy of the Polar Vantage was investigated in a simulated free-living environment at rest and during exercise by comparing the EE estimations of the watch with those of indirect calorimetry. The results revealed a systematic bias of 2 kcal per 10 min (−15 to 12 kcal/10 min) and an MAPE of 21% (9%-31%) during the different activities ranging from sitting and reading to a floorball course. Previous investigations demonstrated higher EE estimation errors of 14% to 210% during walking, running, and sitting [10]; 9% to 43% during lying, sitting, walking, running, and cycling [13]; and 9% to 24% during 13 activities ranging from those of low intensity (eg, writing at a computer) to those of vigorous intensity (eg, elliptical exercise or Wii tennis play) [27]. Shcherbina et al [12] investigated 7 wrist-worn monitors, and the MAPE values in EE estimation ranged from 27% to 93%, depending on the device. As such, these authors claimed that no wrist-worn monitoring devices in 2016 reported EE within an acceptable error range under controlled laboratory conditions during walking, running, and cycling.

Extending from these studies, this study assessed the performance of wrist-worn technologies estimating EE during a larger variety of low-to-high-intensity activities combined with little to a lot of nonsteady wrist and arm movement in a simulated free-living environment. Notably, the present findings are comparable with or better than those found in other recent studies on measurement systems that are also used as reference devices for EE measurement [11,28]. The wearable electrocardiogram Actiheart showed a similar MAPE in EE estimation of 20% (SD 15%), and the temperature- and acceleration-based SenseWear armband showed an MAPE of 39% (SD 18%) in semistructured activities [11,28,29]. To improve EE prediction, individual calibration is needed before each measurement with Actiheart, which hampers its feasibility of use in daily life. In contrast, the findings of this study represent a good overall EE accuracy in the Polar Vantage wristwatch, with the added advantage of its high ease of use.

In this study, larger wrist circumference and higher activity intensity (quantified as perceived exertion) were shown to predict an increased MAE in EE estimation. The highest Borg scale values were reported during the high-intensity activities, demonstrating diminished EE accuracy. This was in contrast to previous studies showing more accurate values during high-intensity activities than during low-to-moderate-intensity activities [10-13]. Noticeably, higher perceived exertion was reported in activities with mainly a lot of wrist and arm movement. Contrary to the fact that higher activity intensity predicted higher MAE, the Polar Vantage showed the very highest error rates during household chores, a low-to-moderate-intensity activity. However, household chores induce much arm movement. This was in line with the results of previous investigations, demonstrating that activities with more wrist and arm movement reveal increased error rates in HR estimation [12,14]. According to the manufacturer, Polar Electro Oy, EE estimation by the Polar Vantage wristwatch is based on the HR measurement and 3D accelerometer signal, with an activity-dependent weighting of these 2 components. During low-intensity activities, more acceleration information is taken into account, and during high-intensity activities, more HR information is taken into account. Moreover, they stated that challenges to accurate HR measurements include the hands facing down, wrist movement, cold skin, and incorrect device placement. Therefore, it is reasonable to state that the EE estimation of the Polar Vantage—and, most likely, many other wrist-worn monitors—is of reduced accuracy in activities that require strong nonsteady wrist and arm movement, regardless of the exercise intensity. Furthermore, EE is dependent on many anthropometric characteristics of the user [30]. In our study, larger wrist circumference revealed an increased MAE in EE estimation, which is in line with previous findings related to wrist-based HR assessments [12,14].

Practical Implications

Generally, the Polar Vantage wristwatch showed promising accuracy in the estimation of EE. However, in activities with strong arm and wrist movements, the EE estimation remains challenging. In some activities, the arm with the mounted monitor takes an active part in the activity, whereas in other activities, it has a passive role. Second, in some activities, the human body is concentrated on doing physical work for a long time in a steady condition, which makes the physiology and calculation of EE more stable than it is in other activities that require a lot of stop and go or require little movement. On the basis of the present findings, the accuracy in EE estimation by the Polar Vantage is activity dependent, and we did not observe a tendency to either under- or overestimate EE. The Polar Vantage recorded EE during the sitting and reading activity—an activity task that is predominantly done during the day—with an acceptable accuracy in 93% of cases. However, the EE during household chores—a compulsory activity for many people and one that is often performed—was poorly assessed. On average, the wristwatch gives a valuable quantification of the training intensity and is a useful indicator of the daily energy output of a person. Nevertheless, such a monitor does not yet give accurate medical guidance or coaching on parameters such as how much one should eat for a balanced energy input and output.

Limitations

The measurements were conducted after the specific exercise modes were selected for each investigated activity, which were ideal setups for testing the monitor. As such, the EE estimations presented in this study were obtained in a training mode and may look different from those measured during a 24×7 assessment of daily life.

Conclusions

This study demonstrated that the multisensory wristwatch Polar Vantage has a statistically moderate-to-good accuracy in EE estimation that is activity dependent. During the sitting and reading activities, the EE estimation is good, whereas during nonsteady activities entailing wrist and arm movement, the EE accuracy is still moderate. However, compared with the other available wrist-worn EE monitors, the Polar Vantage can be

recommended as it performs among the best. To better understand possible inaccurate measurements, users should be aware of the challenges that such technologies must still overcome.

Acknowledgments

Polar Electro Oy (Finland) funded in part the experiment described in this paper.

Authors' Contributions

RGA and TW conceived and designed the research. TS conducted the experiments. RGA analyzed the data and wrote the manuscript. All the authors read and approved the manuscript.

Conflicts of Interest

None declared.

References

1. Blair SN, Horton E, Leon AS, Lee IM, Drinkwater BL, Dishman RK, et al. Physical activity, nutrition, and chronic disease. Med Sci Sports Exerc 1996 Mar;28(3):335-349. [doi: 10.1097/00005768-199603000-00009] [Medline: 8776222]
2. Hamilton MT, Hamilton DG, Zderic TW. Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. Diabetes 2007 Nov;56(11):2655-2667 [FREE Full text] [doi: 10.2337/db07-0882] [Medline: 17827399]
3. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett Jr DR, Tudor-Locke C, et al. 2011 compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
4. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr 2014;1:5 [FREE Full text] [doi: 10.3389/fnut.2014.00005] [Medline: 25988109]
5. Strath SJ, Swartz AM, Bassett Jr DR, O'Brien WL, King GA, Ainsworth BE. Evaluation of heart rate as a method for assessing moderate intensity physical activity. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S465-S470. [doi: 10.1097/00005768-200009001-00005] [Medline: 10993416]
6. Payne PR, Wheeler EF, Salvosa CB. Prediction of daily energy expenditure from average pulse rate. Am J Clin Nutr 1971 Sep;24(9):1164-1170. [doi: 10.1093/ajcn/24.9.1164] [Medline: 5094489]
7. Corder K, Brage S, Wareham NJ, Ekelund U. Comparison of PAEE from combined and separate heart rate and movement models in children. Med Sci Sports Exerc 2005 Oct;37(10):1761-1767. [doi: 10.1249/01.mss.0000176466.78408.cc] [Medline: 16260978]
8. Keytel LR, Goedecke JH, Noakes TD, Hiiloskorpi H, Laukkanen R, van der Merwe L, et al. Prediction of energy expenditure from heart rate monitoring during submaximal exercise. J Sports Sci 2005 Mar;23(3):289-297. [doi: 10.1080/02640410470001730089] [Medline: 15966347]
9. Luke A, Maki KC, Barkey N, Cooper R, McGee D. Simultaneous monitoring of heart rate and motion to assess energy expenditure. Med Sci Sports Exerc 1997 Jan;29(1):144-148. [doi: 10.1097/00005768-199701000-00021] [Medline: 9000168]
10. 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]
11. 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]
12. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):E3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
13. 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]
14. Gilgen-Ammann R, Buller MJ, Bitterle JL, Delves SK, Veenstra BJ, Roos L, et al. Evaluation of pulse rate measurement with a wrist worn device during different tasks and physical activity. Curr Issues Sport Sci 2018;3(11):1-9. [doi: 10.15203/CISS_2018.011]
15. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol 1988 Jun;124(6):869-871. [doi: 10.1001/archderm.124.6.869] [Medline: 3377516]
16. Compher C, Frankenfield D, Keim L, Roth-Yousey L, Evidence Analysis Working Group. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. J Am Diet Assoc 2006 Jun;106(6):881-903. [doi: 10.1016/j.jada.2006.02.009] [Medline: 16720129]

17. Maier T, Gross M, Trösch S, Steiner T, Müller B, Bourban P, et al. Swiss Olympic. 2016. Manual Leistungsdiagnostik URL: https://www.swissolympic.ch/dam/jcr:b15b191a-eb0d-46e8-b9c0-417b887a440d/Leistungsdiagnostik_Manual_160201_DE.pdf [accessed 2019-08-22]
18. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 1970;2(2):92-98. [Medline: 5523831]
19. Horton JF, Stergiou P, Fung TS, Katz L. Comparison of Polar M600 optical heart rate and ECG heart rate during exercise. Med Sci Sports Exerc 2017 Dec;49(12):2600-2607. [doi: 10.1249/MSS.0000000000001388] [Medline: 29135785]
20. Balmer R. MobileSport. 2013. Unihockey Spielend Entdecken URL: https://www.mobilesport.ch/assets/lbwp-cdn/mobilesport/files/2013/12/12_13_Unihockey_d.pdf [accessed 2019-08-22]
21. Gilgen-Ammann R, Schweizer T, Wyss T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. Eur J Appl Physiol 2019 Jul;119(7):1525-1532. [doi: 10.1007/s00421-019-04142-5] [Medline: 31004219]
22. Armstrong LE, Costill DL. Variability of respiration and metabolism: responses to submaximal cycling and running. Res Q Exercise Sport 1985 Jun;56(2):93-96. [doi: 10.1080/02701367.1985.10608441]
23. Vogler AJ, Rice AJ, Gore CJ. Validity and reliability of the Cortex MetaMax3B portable metabolic system. J Sports Sci 2010 May;28(7):733-742. [doi: 10.1080/02640410903582776] [Medline: 20419553]
24. Elia M, Livesey G. Energy expenditure and fuel selection in biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. World Rev Nutr Diet 1992;70:68-131. [doi: 10.1159/000421672] [Medline: 1292242]
25. Jeukendrup AE, Wallis GA. Measurement of substrate oxidation during exercise by means of gas exchange measurements. Int J Sports Med 2005 Feb;26(Suppl 1):S28-S37. [doi: 10.1055/s-2004-830512] [Medline: 15702454]
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [doi: 10.1016/S0140-6736(86)90837-8] [Medline: 2868172]
27. Lee JM, 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]
28. Bhammar DM, Sawyer BJ, Tucker WJ, Lee JM, Gaesser GA. Validity of SenseWear® Armband v5.2 and v2.2 for estimating energy expenditure. J Sports Sci 2016 Oct;34(19):1830-1838 [FREE Full text] [doi: 10.1080/02640414.2016.1140220] [Medline: 26854829]
29. Koehler K, Braun H, de Marées M, Fusch G, Fusch W, Schaenzer W. Assessing energy expenditure in male endurance athletes: validity of the SenseWear Armband. Med Sci Sports Exerc 2011 Jul;43(7):1328-1333. [doi: 10.1249/MSS.0b013e31820750f5] [Medline: 21131865]
30. Zakeri I, Adolph AL, Puyau MR, Vohra FA, Butte NF. Application of cross-sectional time series modeling for the prediction of energy expenditure from heart rate and accelerometry. J Appl Physiol (1985) 2008 Jun;104(6):1665-1673 [FREE Full text] [doi: 10.1152/japplphysiol.01163.2007] [Medline: 18403453]

Abbreviations

EE: energy expenditure
HR: heart rate
HR_{max}: maximal heart rate
MAE: mean absolute error
MAPE: mean absolute percentage error
OR: odds ratio
PAR-Q: physical activity readiness questionnaires
VCO_2: carbon dioxide production
VO_2: oxygen consumption
VO_{2max}: maximal oxygen uptake

Edited by G Eysenbach; submitted 30.04.19; peer-reviewed by T Muto, M Wallen; comments to author 02.07.19; revised version received 15.07.19; accepted 19.07.19; published 02.10.19.

Please cite as:
Gilgen-Ammann R, Schweizer T, Wyss T
Accuracy of the Multisensory Wristwatch Polar Vantage's Estimation of Energy Expenditure in Various Activities: Instrument Validation Study
JMIR Mhealth Uhealth 2019;7(10):e14534
URL: https://mhealth.jmir.org/2019/10/e14534
doi:10.2196/14534
PMID:31579020

Original Paper

Associations Between Heart Rate Variability Measured With a Wrist-Worn Sensor and Older Adults' Physical Function: Observational Study

Sarah Anne Graham[1,2], PhD; Dilip V Jeste[1,2,3], MD; Ellen E Lee[1,2], MD; Tsung-Chin Wu[4], BS; Xin Tu[1,2], PhD; Ho-Cheol Kim[5], PhD; Colin A Depp[1,2,6], PhD

[1]Sam and Rose Stein Institute for Research on Aging, University of California San Diego, La Jolla, CA, United States
[2]Department of Psychiatry, University of California San Diego, La Jolla, CA, United States
[3]Department of Neurosciences, University of California San Diego, La Jolla, CA, United States
[4]Department of Mathematics, University of California San Diego, La Jolla, CA, United States
[5]Scalable Knowledge Intelligence, IBM Research – Almaden, San Jose, CA, United States
[6]VA San Diego Healthcare System, San Diego, CA, United States

Corresponding Author:
Sarah Anne Graham, PhD
Sam and Rose Stein Institute for Research on Aging
University of California San Diego
9500 Gilman Drive, #0664
La Jolla, CA, 92093-0664
United States
Phone: 1 858 534 5433
Email: sagraham@ucsd.edu

Abstract

Background: Heart rate variability (HRV), or variation in beat-to-beat intervals of the heart, is a quantitative measure of autonomic regulation of the cardiovascular system. Low HRV derived from electrocardiogram (ECG) recordings is reported to be related to physical frailty in older adults. Recent advances in wearable technology offer opportunities to more easily integrate monitoring of HRV into regular clinical geriatric health assessments. However, signals obtained from ECG versus wearable photoplethysmography (PPG) devices are different, and a critical first step preceding their widespread use is to determine whether HRV metrics derived from PPG devices also relate to older adults' physical function.

Objective: This study aimed to investigate associations between HRV measured with a wrist-worn PPG device, the Empatica E4 sensor, and validated clinical measures of both objective and self-reported physical function in a cohort of older adults living independently within a continuing care senior housing community. Our primary hypothesis was that lower HRV would be associated with lower physical function. In addition, we expected that HRV would explain a significant proportion of variance in measures of physical health status.

Methods: We evaluated 77 participants from an ongoing study of older adults aged between 65 and 95 years. The assessments encompassed a thorough examination of domains typically included in a geriatric health evaluation. We collected HRV data with the Empatica E4 device and examined bivariate correlations between HRV quantified with the triangular index (HRV TI) and 3 widely used and validated measures of physical functioning—the Short Physical Performance Battery (SPPB), Timed Up and Go (TUG), and Medical Outcomes Study Short Form 36 (SF-36) physical composite scores. We further investigated the additional predictive power of HRV TI on physical health status, as characterized by SF-36 physical composite scores and Cumulative Illness Rating Scale for Geriatrics (CIRS-G) scores, using generalized estimating equation regression analyses with backward elimination.

Results: We observed significant associations of HRV TI with SPPB (n=52; Spearman ρ=0.41; *P*=.003), TUG (n=51; ρ=−0.40; *P*=.004), SF-36 physical composite scores (n=49; ρ=0.37; *P*=.009), and CIRS-G scores (n=52, ρ=−0.43; *P*=.001). In addition, the HRV TI explained a significant proportion of variance in SF-36 physical composite scores (R^2=0.28 vs 0.11 without HRV) and CIRS-G scores (R^2=0.33 vs 0.17 without HRV).

Conclusions: The HRV TI measured with a relatively novel wrist-worn PPG device was related to both objective (SPPB and TUG) and self-reported (SF-36 physical composite) measures of physical function. In addition, the HRV TI explained additional variance in self-reported physical function and cumulative illness severity beyond traditionally measured aspects of physical health. Future steps include longitudinal tracking of changes in both HRV and physical function, which will add important insights regarding the predictive value of HRV as a biomarker of physical health in older adults.

(JMIR Mhealth Uhealth 2019;7(10):e13757) doi:10.2196/13757

KEYWORDS

wearable technology; aging; electrocardiogram; geriatric assessment

Introduction

Background

Heart rate variability (HRV), or variation in beat-to-beat intervals of the heart, is a quantitative measure of autonomic regulation of the cardiovascular system that reflects the ability of the system to react to stressors [1-3]. Low HRV indicates improper coordination between the sympathetic and parasympathetic nervous systems and is a well-established predictor of future cardiovascular disease [4-8]. HRV is also linked to other aspects of health that are directly impacted by autonomic function such as self-regulatory capacity and psychological and physiological stress [5,9,10]. Most HRV metrics are shown to decline normally with age, primarily during younger decades, and some of these metrics increase again after the seventh decade [4,11-15]. In studies of older adults, low HRV measured by electrocardiogram (ECG) recordings has shown preliminary relationships with physical frailty [16-19]. Thus, measurement of HRV may add important information to an assessment of older adults' physical functioning and help to identify individuals at higher risk for physical decline.

Heart Rate Variability Measurement

HRV is traditionally measured in clinical or laboratory settings using standard ECG equipment or in the field with a 24-hour Holter monitor [20]. Although accurate, these measures are time consuming, require expert setup, and are obtrusive, making them less appropriate for health assessments and unlikely to be routine. Recent advances in technology enable measurement of HRV in more ecologically valid settings with unobtrusive wearable devices [21,22]. For example, wrist-worn devices capture HRV via photoplethysmography (PPG) sensors that detect blood volume changes in the microvasculature with each heartbeat [22-24]. These blood volume changes allow the determination of a PPG-derived peak-to-peak (P-P) interval, which is a valid proxy measure of the R-R interval derived from ECG recordings [25-27]. P-P or R-R intervals are also known as interbeat intervals and reflect the durations between successive heartbeats. HRV can be calculated from interbeat intervals with many different metrics (eg, time domain: standard deviation of normal-to-normal index; geometric: triangular index [TI]; and frequency domain: low frequency to high frequency power ratio) [3,20,28]. Selecting the appropriate HRV metric requires careful consideration of the length of recording and quality of the data. Wrist-worn PPG devices for HRV monitoring are increasing in popularity because of their ease of use; however, there is very little literature regarding the clinical use of these devices [29]. A critical step toward understanding the utility of these devices is to determine if and how HRV metrics derived from wrist-worn PPG devices relate to the physical function of older adults. To our knowledge, this is the first study using a wrist-worn PPG device to derive HRV in an older adult population.

Study Purpose

If HRV metrics derived from a wrist-worn wearable are related to clinical measures of physical function and further explain variability in physical health status, a wearable tool could be a useful addition to regular clinical geriatric health assessments for older adults. The primary aim of this study was to investigate associations between HRV measured with a wrist-worn PPG device, the Empatica E4 sensor (Empatica Inc) [30,31], and widely used and validated clinical measures of physical function, including Short Physical Performance Battery (SPPB) scores, Timed Up and Go (TUG) scores [32-34], and self-reported physical function (SF-36 physical composite scores) [35], in a cohort of older adults living independently in a continuing care senior housing community (CCSHC). We selected the SPPB, TUG, and SF-36 as objective and subjective measures of physical capacity and function, respectively, because they are a few of the most widely used measures of physical function in clinical geriatric research and health assessments [36-39]. Our primary hypothesis was that lower HRV would be associated with lower physical function. We further investigated whether HRV could explain additional variability in physical health status, as measured by the SF-36 physical composite scores and the presence and severity of physical comorbidities via the Cumulative Illness Severity Scale for Geriatrics (CIRS-G) [40], beyond traditionally documented aspects of health such as age, gender, race, blood pressure, medication and alcohol use, smoking status, and anthropometric measurements.

Methods

Participants

The University of California San Diego (UCSD) Human Research Protections Program (HRPP) approved the study protocol. Research staff recruited participants from a CCSHC in San Diego County via short presentations using an HRPP-approved script and flyers. We recruited 77 participants living in the independent living sector of the CCSHC, and all of them provided written informed consent before study participation. The inclusion and exclusion criteria for enrollment were (1) English-speaking individuals older than 65 years, (2) ability to complete study assessments, and (3) no known diagnosis of dementia or other disabling illness.

Procedures

We evaluated participants during the baseline assessment in an ongoing study of older adults between 65 and 95 years of age [41]. This assessment encompassed a thorough examination of domains typically included in a geriatric health evaluation: sociodemographic and medical health information, physical function measurements, cognitive measurements, and additional assessments of everyday function. The mean duration of the assessments was 2:34:31 (HH:MM:SS; range 1:09:30-4:38:10). The range of activities performed during this assessment reflected the types of tasks that an individual may encounter in their daily lives, and thus, the measure of HRV obtained from this time period is more similar to HRV metrics obtained over longer durations (eg, 24 hours), reflecting the cardiovascular system's response to a range of environmental stimuli and workloads, as opposed to short-term measurements (eg, <5 min) that reflect immediate responses to a particular stimulus [3].

For this investigation, we characterized participants with sociodemographic and clinical information including age, sex, race, education, body mass index (BMI), waist-to-hip ratio, blood pressure, medication use (antihypertensives and antidepressants), presence and severity of physical comorbidities reported with the CIRS-G [40], and capacity for everyday functioning with the UCSD Performance-Based Skills Assessment Brief [42].

Physical Function Measurement

We assessed physical functioning with both objective, capacity-based measures and subjective, self-reported measures. Capacity-based measures included the SPPB and TUG test, both of which have been shown to be valid, objective measures of lower extremity function and mobility [32-34]. The SPPB has 3 subcategories: the ability to stand for 10 seconds with feet in 3 different positions (side by side, semitandem, and tandem), 2 timed trials of a 3- or 4-m walk (faster of the two), and the time it takes to rise from a chair 5 times. Scores range from 0 to 12, with higher scores indicating better lower extremity function. The TUG test is scored by the amount of time it takes to rise from a chair, walk 3 m at a comfortable pace, turn, return to the chair, and sit down. A shorter time to complete the test indicates better mobility. For self-reported physical functioning, we used the physical composite score from the Medical Outcomes Study Short Form 36 (SF-36), which is a globally used questionnaire for assessing 8 dimensions of health-related quality of life [35]. The physical composite score is an aggregate of the 8 scale scores reflecting self-reported physical health [43].

Heart Rate Variability Measurement

We collected raw PPG signals and interbeat intervals for the calculation of HRV with a wrist-worn device called the Empatica E4 [30,31]. Research staff placed the Empatica E4 wristband on a participant's nondominant wrist at the onset of the assessment. During the first few minutes, the staff member checked the quality of the PPG signal in real time through the E4 application to ensure proper wrist placement as recommended by Empatica. Wrist-worn PPG sensors tend to be more accurate at rest than during exercise because of contamination from movement artifacts and often require accelerometry technology to measure consistent or repetitive movements to minimize influence of these artifacts [21,44-47]. The PPG sensor in the Empatica E4 is designed to be robust against movement artifacts in that it can attenuate noise even when movements are not repetitive in nature, using an artifact removal technique based on a combination of multiple infrared light wavelengths [30].

The Empatica E4 provided continuous heart rate and interbeat intervals that were not interpolated. It removed interbeat intervals that corresponded to regions where the PPG signal was not clear. For postprocessing, we downloaded the interbeat intervals provided in a comma-separated value file format from the E4 connect application. We viewed each interbeat interval time series in MATLAB R2016a (MathWorks, Inc) and calculated the percentage of gaps in the data (ie, nonconsecutive interbeat intervals) for each participant. We removed 19 participants from data analyses because of poor-quality interbeat interval recordings with greater than 20% gaps between interbeat intervals that resulted in less than 20 min of recorded high-quality interbeat interval data. We then processed the remaining data with Kubios HRV standard software version 3.1.0 using a low threshold artifact correction to adjust for any remaining ectopic beats and the smoothing priors detrending method (default λ=500) [48,49].

Selecting an HRV metric requires careful consideration of the strengths and limitations of wrist-worn PPG devices. The Empatica E4 uses algorithms to remove errant interbeat intervals and provides the remaining *clean* data that contain regions where interbeat intervals are not necessarily consecutive [30,31,50]. To overcome this limitation and provide a robust and valid HRV metric, we selected a metric that is less sensitive to gaps between interbeat intervals called the HRV TI [44,51,52]. The HRV TI is a geometric index calculated as the integral of the density distribution of interbeat intervals divided by the maximum of the density distribution [28] with larger numbers indicating more favorable HRV. The HRV TI requires a longer recording period (approximately 20 min), is robust against missing interbeat intervals, and has good intraindividual reproducibility [28]. For these reasons, we selected the HRV TI and analyzed the longest recording period available for each participant (ie, full-length assessment) for greater measurement stability.

Statistical Analyses

Given the exploratory nature of this study, we did not perform a formal a priori power analysis. We performed correlational statistical analyses in SPSS version 25 and generalized estimating equation (GEE) regression with backward elimination in R version 3.4.1. We used Spearman correlation coefficients to assess bivariate correlations between the HRV TI measured across the full assessment period and SPPB, TUG, and SF-36 physical composite scores. We removed an additional 5 participants from correlational analyses for the suspected presence of arrythmia based on an HRV TI greater than 20.42 [53]. We were also missing SPPB scores for 1 participant, TUG scores for 2 participants, and SF-36 physical composite scores for 4 participants. We included the remaining participants in analyses (n=52, SPPB; n=51, TUG; and n=49, SF-36 physical

composite). We conducted Mann-Whitney U tests to determine if there were differences between the participants removed from analyses and those retained. For our primary analyses, we first examined correlations between the HRV TI and variables that have been previously shown to be related to HRV including age, BMI, blood pressure, and mean heart rate [4,54]. We also looked for group differences in the HRV TI between men and women and between participants on versus off antihypertensive and antidepressant medications [16]. We used these results to determine whether to adjust for any of these variables using partial correlations (ie, control for one or more of these variables using nonparametric partial correlation syntax in SPSS if any covariates could potentially modify our primary bivariate relationships of interest). For the regression analyses, we included n=42 participants with complete data on all potential covariates (age, gender, race, blood pressure, BMI, waist-to-hip ratio, SPPB and TUG scores, antihypertensive and antidepressant use, alcohol use, smoking status, SF-36 physical composite or CIRS-G scores depending on the model, and HRV). We checked variance inflation factors (VIFs) for each covariate and considered a covariate with VIF greater than 3 as having high multicollinearity [55]. We rebuilt linear models and recalculated VIFs for each covariate after excluding covariates with a VIF greater than 3 and repeated this exclusion process until all covariates had a VIF less than 3. To build the regression model, we started with all variables of VIF less than or equal to 3 in the GEE model [56]. We then iteratively removed the variable with the largest *P* value, rebuilt a new GEE model based on the remaining variables, and recalculated *P* values for each variable. We repeated this process until all remaining variables had a *P* value at a threshold of less than or equal to .2. The backward elimination procedure ensured minimum bias in the final model [57]. We derived R^2 values using the linear model. We established statistical threshold for each family of statistical tests: correlational significance to a Bonferroni adjusted $P \leq .017$ (0.05/3 primary outcomes of interest) and $P \leq .025$ for the regressions (0.05/2 regression analyses).

Results

We did not find significant differences for any variable in Table 1 between the individuals removed from correlational analyses and those that we retained (P=.17 to P=.99). Heart rate remained within a normal range during the baseline assessment: mean of 76 beats per minute (range 63-93). Mean heart rate over the course of the assessment was inversely correlated with the mean interbeat interval length $\rho = -0.78$ and $P \leq .001$ (Figure 1) as expected, because a higher heart rate corresponds to a shorter duration between beats and vice versa.

There were no significant differences in the HRV TI based on sex (U=242.5; P=.40), antihypertensive medication use (U=271.5; P=.37), or antidepressant medication use (U=147.0; P=.59). We also did not observe significant correlations between any potential covariate and both HRV TI and physical function measures (ie, variables that could potentially modify these bivariate relationships; Table 2). Therefore, we did not adjust for any variables using partial correlations for our primary hypothesis regarding relationships between HRV TI and physical function measures. HRV TI was significantly related to SPPB, TUG, and SF-36 physical composite scores (Table 2 and Figure 2). We also observed an inverse relationship between HRV TI and CIRS-G scores (Figure 2).

Results of the regression revealed that the HRV TI explained a significant proportion of variance in physical health status as characterized by both SF-36 physical composite and CIRS-G scores. R^2 values increased from 0.11 to 0.28, with the HRV TI included in estimating SF-36 physical composite scores, and from 0.17 to 0.33, with the HRV TI included in estimating CIRS-G scores. Additional significant correlates for these measures of physical health status following the backward elimination procedure included gender, BMI, medication use, and smoking status (Table 3). VIF values were all less than 2.0, suggesting minimal collinearity.

Table 1. Sociodemographic information and other variables collected from participants.

Variable name	Values
Age (years; n=77), mean (SD); range	82.9 (6.7); 67-98
Gender (n=77); number of females, n (%)	52 (68)
Education (years; n=77), mean (SD); range	15.8 (2.4); 12-20
Race (n=77); number of whites, n (%)	68 (88)
Body mass index (kg/m^2; n=76), mean (SD); range	27.9 (4.9); 19-43
Waist-to-hip ratio (ratio; n=75), mean (SD); range	0.87 (0.08); 0.71-1.06
Systolic blood pressure (mm Hg; n=74), mean (SD); range	134 (17); 100-167
Diastolic blood pressure (mm Hg; n=74), mean (SD); range	75 (9); 56-94
University of California San Diego Performance-Based Skills Assessment Brief (score; n=77), mean (SD); range	76.1 (12.5); 37.4-100.0
Cumulative Illness Rating Scale for Geriatrics (score; n=76), mean (SD); range	8.8 (3.2); 2-15
Antihypertensives (n=75); number of participants using, n (%)	48 (62)
Antidepressants (n=75); number of participants using, n (%)	13 (17)
Current smoker (n=75); number of participants smoking, n (%)	1 (1)
Alcohol use (n=68), n (%)	
Lifetime abstainer (number of yes; [% abstaining])	6 (9)
Current infrequent drinker (number of yes; [% drinking])	36 (53)
Current regular drinker (number of yes; [% drinking])	17 (25)
Former drinker (number of yes; [% used to drink])	9 (13)
Short Physical Performance Battery (score out of 12; n=76), mean (SD); range	8.2 (2.7); 0-12
Timed up and Go time (seconds; n=74), mean (SD; range)	11.0 (3.1); 6.1-23.0
Short Form 36 physical composite (score out of 100; n=70), mean (SD); range	41.8 (10.7); 18.3-60.1
Heart rate variability triangular index (n=53), mean (SD); range	11.4 (2.9); 3.7-17.2

Figure 1. Scatterplot of the relationship between mean heart rate across the entire evaluation and mean interbeat interval length: n=52; Spearman correlation=−0.78; P≤.001. bpm: beats per minute.

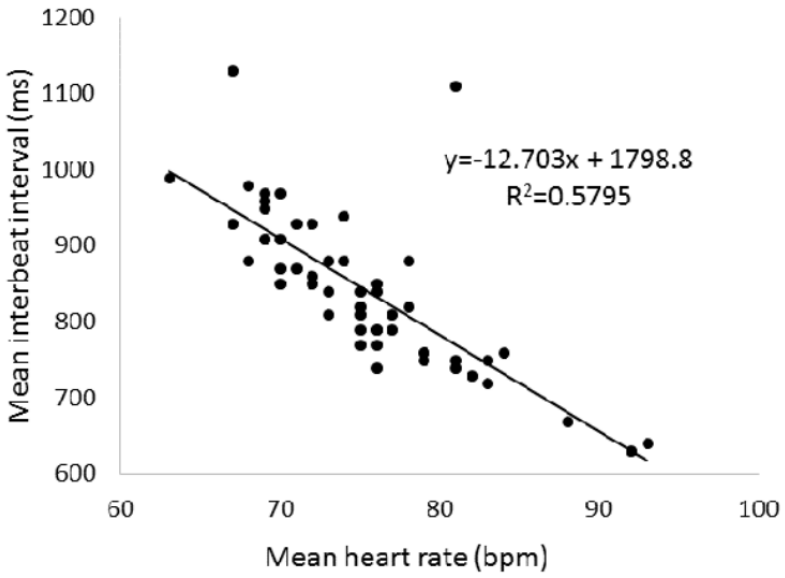


Table 2. Bivariate correlations among all variables. Correlations for primary hypotheses are in italics.

Variables	ρ (*P* value)									
	SPPB[a] (n=52)	TUG[b] (n=51)	SF-36 phys comp[c] (n=49)	CIRS-G[d] (n=52)	Age (n=53)	BMI[e] (n=51)	Waist to hip ratio (n=52)	Sys BP[f] (n=50)	Dia BP[g] (n=50)	Mean HR[h] (n=53)
HRV TI[i] (n=53)	*0.41*[j]; (*.003*)	*−0.40*[j]; (*.004*)	*0.37*[j]; (*.009*)	−0.43[j]; (.001)	−0.10; (.47)	0.07; (.64)	−0.20; (.16)	0.07; (.63)	0.01; (.97)	−0.36[j]; (.009)
SPPB	—[k]	−0.79[l]; (≤.001)	0.59[l]; (≤.001)	−0.23; (.11)	−0.29; (.04)	0.10; (.48)	−0.08; (.58)	0.06; (.69)	0.10; (.49)	−0.04; (.80)
TUG	—	—	−0.39[j]; (.007)	0.19; (.18)	0.40[j]; (.004)	−0.15; (.30)	0.10; (.50)	−0.02; (.87)	−0.13; (.39)	0.18; (.20)
SF-36 phys comp	—	—	—	−0.65[l]; (≤.001)	−0.24; (.10)	0.07; (.65)	−0.08; (.59)	−0.08; (.61)	−0.10; (.95)	−0.07; (.65)
CIRS-G	—	—	—	—	−0.20; (.16)	0.19; (.19)	−0.05; (.75)	−0.02; (.88)	−0.03; (.84)	0.09; (.54)
Age	—	—	—	—	—	−0.14; (.34)	0.40[l]; (.003)	0.35[j]; (.01)	0.18; (.21)	0.00; (>.99)
BMI	—	—	—	—	—	—	0.01; (.95)	−0.12; (.42)	0.13; (.37)	−0.21; (.14)
Waist to hip ratio	—	—	—	—	—	—	—	0.13; (.36)	0.05; (.74)	−0.09; (.51)
Sys BP	—	—	—	—	—	—	—	—	0.64[l]; (≤.001)	−0.11; (.46)
Dia BP	—	—	—	—	—	—	—	—	—	−0.03; (.85)

[a]SPPB: Short Physical Performance Battery.

[b]TUG: Timed Up and Go.

[c]SF-36 phys comp: Short Form 36 physical composite score.

[d]CIRS-G: Cumulative Illness Rating Scale for Geriatrics.

[e]BMI: body mass index.

[f]Sys BP: systolic blood pressure.

[g]Dia BP: diastolic blood pressure.

[h]HR: heart rate.

[i]HRV TI: heart rate variability triangular index.

[j]*P*≤.017.

[k]Not applicable.

[l]*P*≤.001.

Figure 2. Scatter plots of the relationships between heart rate variability triangular index (TI) and (a) Short Physical Performance Battery scores (n=52), (b) Timed Up and Go scores (n=51), (c) Short Form 36 physical composite scores (n=49), and (d) Cumulative Illness Rating Scale for Geriatrics (CIRS-G) scores (n=52). HRV: heart rate variability; SF-36: Short Form 36.

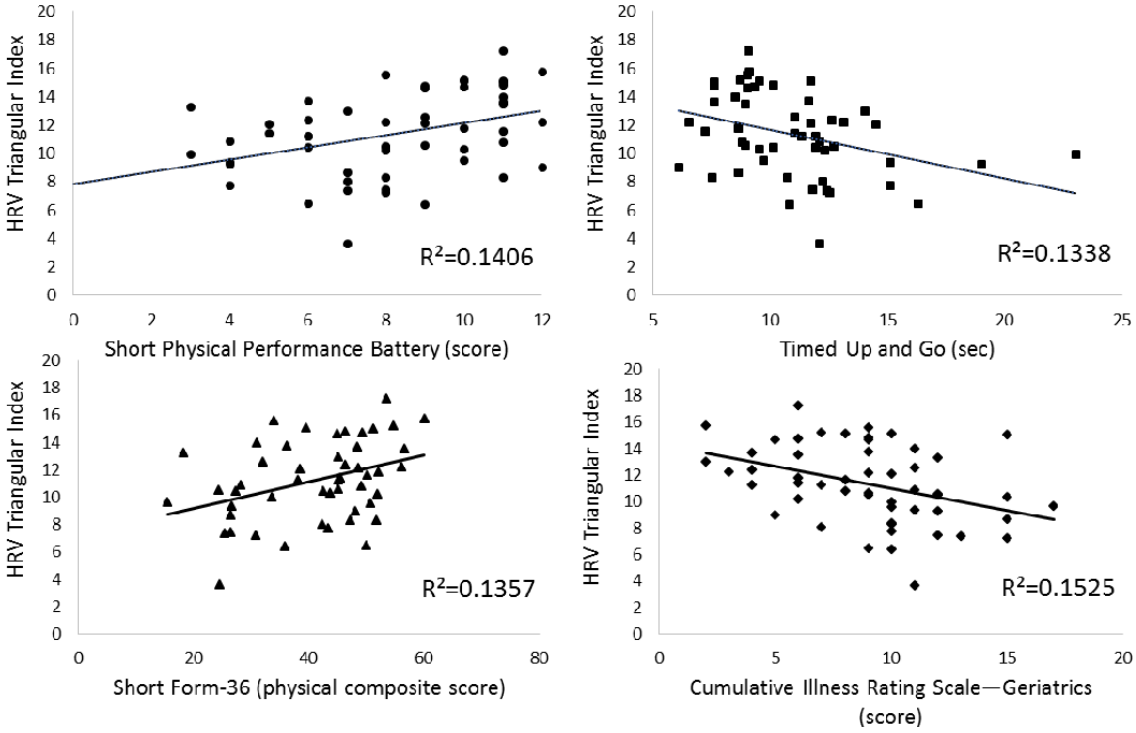


Table 3. Results of the regression analyses (backward selection): remaining significant correlates of physical health status characterized by Short Form 36 physical composite scores and Cumulative Illness Rating Scale for Geriatrics scores. Adjusted R^2 of the full model for SF-36 scores=0.28 and for CIRS-G scores=0.33.

Parameter	Estimate	Naïve SE	Naïve z	Robust SE	Robust z	*P* value
SF-36[a] physical composite score						
Intercept	22.01	5.81	3.79	6.05	3.64	≤.001
HRV TI[b]	1.58	0.47	3.38	0.44	3.58	≤.001
Gender	5.88	3.01	1.96	2.79	2.11	.04
Antihypertensives	−5.95	2.71	−2.20	2.53	−2.35	.02
Smoking status	6.72	2.71	2.48	2.46	2.73	.006
CIRS-G[c] score						
Intercept	8.62	3.02	2.85	2.80	3.07	.002
HRV TI[d]	−0.46	0.14	−3.41	0.10	−4.56	≤.001
Gender	−1.56	0.89	−1.75	0.81	−1.93	.05
BMI[e]	0.15	0.09	1.70	0.07	1.96	.05
Antihypertensives	2.52	0.83	3.05	0.71	3.57	≤.001
Antidepressants	1.34	0.91	1.47	0.95	1.41	.16

[a]SF-36: short form 36.

[b]HRV TI: heart rate variability triangular index

[c]CIRS-G: Cumulative Illness Rating Scale for Geriatrics.

[d]HRV: heart rate variability.

[e]BMI: body mass index.

Discussion

Principal Findings

In this study, we explored relationships between HRV measured using a wrist-worn device and both capacity-based and self-reported measures of physical function obtained during a standard clinical geriatric assessment of older adults living independently in a CCSHC. Our primary hypothesis was supported as we observed significant relationships between the HRV TI and both capacity-based measures of physical function—that is, the SPPB and TUG [32-34] and self-reported physical function, as measured by the SF-36 physical composite score [35,43]. In addition, we observed that the HRV TI explained a significant proportion of variance in physical health status, as characterized by SF-36 physical composite and CIRS-G scores [40].

Relationships Between Heart Rate Variability, Physical Function, and Physical Health

The relationships between the HRV TI and both the SPPB and TUG and self-reported measures of physical function are noteworthy. Self-reported and capacity-based measures of physical capacity are reported to contribute information about physical function in different ways [58]. For example, self-reported measures are more effective in distinguishing among persons at lower levels of physical function who may not be able to complete capacity-based tests [59]. In contrast, self-reported measures fail to differentiate well among persons in the mid- or high range of functioning. Capacity-based measures provide information across a broader range of physical function and discriminate more effectively than self-reported measures at higher capacities [59]. We observed significant relationships between SPPB and TUG scores and self-reported physical function in our sample. We also observed an inverse relationship between the presence of physical comorbidities and self-reported physical function, which is consistent with previous reports demonstrating that older adults who report more health problems also report lower levels of perceived function [60]. Presence of physical comorbidities (CIRS-G) was related to the HRV TI but not to either SPPB or TUG scores in this study. Therefore, HRV TI may characterize a complementary aspect of older adults' physical health that is not captured by physical capacity measures.

Previous work has identified relationships between HRV quantified via standard ECG or Holter monitoring and aspects of older adults' physical function [5,16,19]. In particular, frailty appears to be linked to a loss of physiologic flexibility, as quantified by nonlinear HRV dynamics [16,17]. If changes in HRV precede physical decline, then regular measurement of HRV may be a useful biomarker for older adults. Low HRV, along with male sex, older age, and smoking, has also been reported to be a significant predictor of chronic diseases including hypertension and hyperglycemia and a diagnosis of diabetes within 12 years [61]. These findings suggest that monitoring of HRV may help to identify individuals at high risk for both physical decline and development of chronic diseases.

High HRV has been described as a measure of adaptability [62], which is directly related to maintenance of physical function. The consistent relationships observed between the HRV TI and measures of physical function and physical health in this study suggest that HRV TI quantified with a wearable device is indeed related to physical function and cumulative illness burden in older adults. These relationships were not accounted for by a direct association between HRV and age, suggesting that this measure of autonomic function may be useful for aging research. This finding is not surprising given that previous relationships between HRV and age have generally been observed across a much broader age span [4,14,15,28]. Although standard ECG and Holter monitor recordings have been useful for risk stratification in a variety of pathologic conditions [62-64], our findings suggest that a wearable device may also be useful for quantifying autonomic dysfunction in older adults. Although HRV was significantly associated with physical capacity measures, mean heart rate was not, suggesting that measurement of heart rate alone does not capture the relationship between autonomic regulation and physical function.

Wearable Technology for Heart Rate Variability Measurement

The accessibility of wearable technology [29] and the relationships between HRV derived from a wrist-worn wearable and measures of physical function suggest that incorporating wearables for HRV monitoring into clinical geriatric assessments could potentially help in the early detection of physical decline. In particular, watch-like devices are ideal because they are noninvasive and small and do not interfere with the activities performed during clinical geriatric health assessments. These devices are also relatively low cost, do not require expert setup, and may be more easily deployed than ECG equipment in remote home-health settings. In the future, such wearables may be ideal for monitoring the health of older adults, further enabling early detection of physical decline and early intervention to improve autonomic regulation and possibly delay deterioration of physical function. However, such long-term monitoring of HRV via wearable devices will necessitate careful consideration of the adoption of technology by older adults and ethical concerns such as data privacy and informed consent [65]. It may also be important to monitor HRV in younger populations, given the observed relationships to measures of physical health status in this study and the fact that declining physical function does not happen only in adults older than 65 years. HRV is known to decline with age, particularly before the age of 60 years, but these correlations are only modest [14]. Thus, perhaps those with poorer physical health also exhibit greater tendency for decline. If HRV metrics are related to the physical health status of younger adults as well, early identification of declining HRV may enable more timely intervention to mitigate further decline.

Future Goals for Heart Rate Variability Measurement

In this study, we measured HRV during a single assessment to demonstrate relationships between HRV metrics quantified with a wearable device and clinical measures of physical function; however, we intend to track this cohort of participants longitudinally. These findings will help determine whether HRV

from a wearable device predicts or coincides with changes in physical function. We are also interested in the feasibility of more frequent HRV measurements using wrist-worn devices for older adults and plan to examine day-to-day variability of HRV metrics. The clinical value of HRV measurement is likely to be at the level of the individual and not the group; thus, identifying changes outside of an individual's normal HRV range may lend the most insight into incipient decline.

Study Limitations

This study had several limitations. This was a small sample of predominantly white and middle-to-upper class older adults, so our findings may not apply to other populations. We also had a majority of female participants, which may have prevented us from detecting differences in HRV due to gender. However, gender differences in HRV have been reported to disappear after the age of 50 years [14]. Furthermore, the cross-sectional nature of this investigation precludes the ability to draw causal relationships between the variables investigated. We investigated only 1 PPG device, the Empatica E4 sensor, in this study. However, we did not use an HRV metric provided by this device (Empatica does not offer these data). We simply used the device to collect raw interbeat interval data and then used the well-known Kubios HRV software [48] to calculate HRV TI. The presence of missing interbeat intervals influenced our choice of HRV metric (we selected the HRV TI instead of others) and, in particular, prevented us from using frequency-based HRV metrics as they are affected by missing interbeat intervals. As frequency-based metrics are reported to be more descriptive of the balance between sympathetic and parasympathetic activity [66], improving the quality of interbeat interval recordings from wearable devices is of considerable importance. We also had to remove a considerable number of participants from statistical analyses because of poor-quality interbeat interval recordings. Future research directed at refining wearable technology such as the Empatica E4 to improve the quality of recordings (eg, further minimize movement artifacts) will be essential in improving the utility of these devices. Finally, we cannot easily compare the HRV TI values obtained in this study to previously established norms for several reasons, including less frequent reporting of the HRV TI metric, differences in calculation methods, and differing durations of recordings (eg, often done over a 24-hour period).

Conclusions

In summary, we demonstrated that the HRV TI measured using a relatively novel wearable device, the Empatica E4 sensor, was related to both objective (SPPB and TUG) and self-reported (SF-36 physical composite) measures of physical function of older adults and cumulative illness burden (CIRS-G) collected during a single assessment. In addition, the HRV TI explained a significant proportion of variance in physical health status, as characterized by the SF-36 physical composite scores and CIRS-G scores, beyond typically measured aspects of physical health. The next steps include longitudinal tracking of changes in both HRV and physical function, which will add important insights regarding the possible predictive value of HRV as a biomarker of functional outcomes in older adults.

Acknowledgments

The authors thank all the study participants for their contributions to this work. The funding for this study was provided, in part, by the National Institute of Mental Health T32 Geriatric Mental Health Program (grant MH019934 to Dilip V Jeste [Principal Investigator]), the Stein Institute for Research on Aging at UCSD, and by IBM Research AI through the AI Horizons Network IBM-UCSD AI for Healthy Living program.

Conflicts of Interest

None declared.

References

1. Berntson GG, Bigger JT, Eckberg DL, Grossman P, Kaufmann PG, Malik M, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology 1997 Nov;34(6):623-648. [doi: 10.1111/j.1469-8986.1997.tb02140.x] [Medline: 9401419]
2. Ernst G. Heart-rate variability-more than heart beats? Front Public Health 2017;5:240 [FREE Full text] [doi: 10.3389/fpubh.2017.00240] [Medline: 28955705]
3. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
4. Almeida-Santos MA, Barreto-Filho JA, Oliveira JL, Reis FP, da Cunha OC, Sousa AC. Aging, heart rate variability and patterns of autonomic regulation of the heart. Arch Gerontol Geriatr 2016;63:1-8. [doi: 10.1016/j.archger.2015.11.011] [Medline: 26791165]
5. McCraty R, Shaffer F. Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. Glob Adv Health Med 2015 Jan;4(1):46-61 [FREE Full text] [doi: 10.7453/gahmj.2014.073] [Medline: 25694852]
6. Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol 2010 May 28;141(2):122-131. [doi: 10.1016/j.ijcard.2009.09.543] [Medline: 19910061]

7. Fei LU, Keeling PJ, Sadoul N, Copie X, Malik M, McKenna WJ, et al. Decreased heart rate variability in patients with congestive heart failure and chronotropic incompetence. Pacing Clin Electrophysiol 1996 Apr;19(4 Pt 1):477-483. [doi: 10.1111/j.1540-8159.1996.tb06519.x] [Medline: 8848396]
8. Kleiger RE, Stein PK, Bigger JT. Heart rate variability: measurement and clinical utility. Ann Noninvasive Electrocardiol 2005 Jan;10(1):88-101. [doi: 10.1111/j.1542-474X.2005.10101.x] [Medline: 15649244]
9. Kemp AH, Quintana DS. The relationship between mental and physical health: insights from the study of heart rate variability. Int J Psychophysiol 2013 Sep;89(3):288-296. [doi: 10.1016/j.ijpsycho.2013.06.018] [Medline: 23797149]
10. Williams DP, Cash C, Rankin C, Bernardi A, Koenig J, Thayer JF. Resting heart rate variability predicts self-reported difficulties in emotion regulation: a focus on different facets of emotion regulation. Front Psychol 2015;6:261 [FREE Full text] [doi: 10.3389/fpsyg.2015.00261] [Medline: 25806017]
11. Jandackova VK, Scholes S, Britton A, Steptoe A. Are changes in heart rate variability in middle-aged and older people normative or caused by pathological conditions? Findings from a large population-based longitudinal cohort study. J Am Heart Assoc 2016 Feb 12;5(2):e002365 [FREE Full text] [doi: 10.1161/JAHA.115.002365] [Medline: 26873682]
12. Reardon M, Malik M. Changes in heart rate variability with age. Pacing Clin Electrophysiol 1996 Nov;19(11 Pt 2):1863-1866. [doi: 10.1111/j.1540-8159.1996.tb03241.x] [Medline: 8945057]
13. Santillo E, Migale M, Fallavollita L, Marini L, Balestrini F. Electrocardiographic analysis of heart rate variability in aging heart. In: Advances in Electrocardiograms - Methods and Analysis. Cambridge, England: Open Access Books; 2012.
14. Umetani K, Singer DH, McCraty R, Atkinson M. Twenty-four hour time domain heart rate variability and heart rate: relations to age and gender over nine decades. J Am Coll Cardiol 1998 Mar 1;31(3):593-601 [FREE Full text] [doi: 10.1016/s0735-1097(97)00554-8] [Medline: 9502641]
15. O'Brien IA, O'Hare P, Corrall RJ. Heart rate variability in healthy subjects: effect of age and the derivation of normal ranges for tests of autonomic function. Br Heart J 1986 Apr;55(4):348-354 [FREE Full text] [doi: 10.1136/hrt.55.4.348] [Medline: 3964501]
16. Chaves PH, Varadhan R, Lipsitz LA, Stein PK, Windham BG, Tian J, et al. Physiological complexity underlying heart rate dynamics and frailty status in community-dwelling older women. J Am Geriatr Soc 2008 Sep;56(9):1698-1703 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.01858.x] [Medline: 19166446]
17. Katayama PL, Dias DP, Silva LE, Virtuoso-Junior JS, Marocolo M. Cardiac autonomic modulation in non-frail, pre-frail and frail elderly women: a pilot study. Aging Clin Exp Res 2015 Oct;27(5):621-629. [doi: 10.1007/s40520-015-0320-9] [Medline: 25673231]
18. Koopman JJ, van Bodegom D, Maan AC, Li Z, Ziem JB, Westendorp RG, et al. Heart rate variability, but not heart rate, is associated with handgrip strength and mortality in older Africans at very low cardiovascular risk: a population-based study. Int J Cardiol 2015;187:559-561 [FREE Full text] [doi: 10.1016/j.ijcard.2015.03.383] [Medline: 25863303]
19. Varadhan R, Chaves PH, Lipsitz LA, Stein PK, Tian J, Windham BG, et al. Frailty and impaired cardiac autonomic control: new insights from principal components aggregation of traditional heart rate variability indices. J Gerontol A Biol Sci Med Sci 2009 Jun;64(6):682-687 [FREE Full text] [doi: 10.1093/gerona/glp013] [Medline: 19223607]
20. . Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J 1996 Mar;17(3):354-381. [Medline: 8737210]
21. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]
22. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018;9:743 [FREE Full text] [doi: 10.3389/fphys.2018.00743] [Medline: 30002629]
23. Kamišalić A, Fister I, Turkanović M, Karakatič S. Sensors and functionalities of non-invasive wrist-wearable devices: a review. Sensors (Basel) 2018 May 25;18(6):E1714 [FREE Full text] [doi: 10.3390/s18061714] [Medline: 29799504]
24. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable photoplethysmographic sensors—past and present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]
25. Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can wearable devices accurately measure heart rate variability? A systematic review. Folia Med (Plovdiv) 2018 Mar 1;60(1):7-20. [doi: 10.2478/folmed-2018-0012] [Medline: 29668452]
26. Gil E, Orini M, Bailón R, Vergara JM, Mainardi L, Laguna P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. Physiol Meas 2010 Sep;31(9):1271-1290. [doi: 10.1088/0967-3334/31/9/015] [Medline: 20702919]
27. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 2013 Jun 5;166(1):15-29. [doi: 10.1016/j.ijcard.2012.03.119] [Medline: 22809539]
28. Ziegler D, Piolot R, Strassburger K, Lambeck H, Dannehl K. Normal ranges and reproducibility of statistical, geometric, frequency domain, and non-linear measures of 24-hour heart rate variability. Horm Metab Res 1999 Dec;31(12):672-679. [doi: 10.1055/s-2007-978819] [Medline: 10668921]

29. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202 [FREE Full text] [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
30. Garbarino M, Lai M, Bender D, Picard RW, Tognetti S. Empatica E3 — A Wearable Wireless Multi-Sensor Device for Real-Time Computerized Biofeedback and Data Acquisition. In: Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies. 2014 Presented at: MOBIHEALTH'14; November 3-5, 2014; Athens, Greece p. 39-42. [doi: 10.1109/MOBIHEALTH.2014.7015904]
31. Mccarthy C, Pradhan N, Redpath C, Adler A. Validation of the Empatica E4 Wristband. In: Proceedings of the EMBS International Student Conference. 2016 Presented at: ISC'16; May 29-31, 2016; Ottawa, ON, Canada p. 1-4. [doi: 10.1109/EMBSISC.2016.7508621]
32. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
33. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci 2000 Apr;55(4):M221-M231. [doi: 10.1093/gerona/55.4.m221] [Medline: 10811152]
34. Podsiadlo D, Richardson S. The timed 'Up & Go': a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
35. McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993 Mar;31(3):247-263. [doi: 10.1097/00005650-199303000-00006] [Medline: 8450681]
36. Ward KT, Reuben DB. UpToDate. 2003. Comprehensive Geriatric Assessment. URL: https://www.uptodate.com/contents/comprehensive-geriatric-assessment [accessed 2019-01-15]
37. Roedl KJ, Wilson LS, Fine J. A systematic review and comparison of functional assessments of community-dwelling elderly patients. J Am Assoc Nurse Pract 2016 Mar;28(3):160-169. [doi: 10.1002/2327-6924.12273] [Medline: 25964008]
38. Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, et al. Physical performance measures in the clinical setting. J Am Geriatr Soc 2003 Mar;51(3):314-322. [doi: 10.1046/j.1532-5415.2003.51104.x] [Medline: 12588574]
39. Freiberger E, de Vreede P, Schoene D, Rydwik E, Mueller V, Frändin K, et al. Performance-based physical function in older community-dwelling persons: a systematic review of instruments. Age Ageing 2012 Nov;41(6):712-721. [doi: 10.1093/ageing/afs099] [Medline: 22885845]
40. Parmelee PA, Thuras PD, Katz IR, Lawton MP. Validation of the cumulative illness rating scale in a geriatric residential population. J Am Geriatr Soc 1995 Feb;43(2):130-137. [doi: 10.1111/j.1532-5415.1995.tb06377.x] [Medline: 7836636]
41. Jeste DV, Glorioso D, Lee EE, Daly R, Graham S, Liu J, et al. Study of independent living residents of a continuing care senior housing community: sociodemographic and clinical associations of cognitive, physical, and mental health. Am J Geriatr Psychiatry 2019 Sep;27(9):895-907. [doi: 10.1016/j.jagp.2019.04.002] [Medline: 31078382]
42. Mausbach BT, Harvey PD, Goldman SR, Jeste DV, Patterson TL. Development of a brief scale of everyday functioning in persons with serious mental illness. Schizophr Bull 2007 Nov;33(6):1364-1372 [FREE Full text] [doi: 10.1093/schbul/sbm014] [Medline: 17341468]
43. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-item health survey 1.0. Health Econ 1993 Oct;2(3):217-227. [doi: 10.1002/hec.4730020305] [Medline: 8275167]
44. Baek HJ, Shin JW. Effect of missing inter-beat interval data on heart rate variability analysis using wrist-worn wearables. J Med Syst 2017 Aug 15;41(10):147. [doi: 10.1007/s10916-017-0796-2] [Medline: 28812280]
45. Kos M, Li X, Khaghani-Far I, Gordon CM, Pavel M, Jimison HB. Can accelerometry data improve estimates of heart rate variability from wrist pulse PPG sensors? Conf Proc IEEE Eng Med Biol Soc 2017 Jul;2017:1587-1590 [FREE Full text] [doi: 10.1109/EMBC.2017.8037141] [Medline: 29060185]
46. Zhang Z, Pi Z, Liu B. TROIKA: a general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. IEEE Trans Biomed Eng 2015 Feb;62(2):522-531. [doi: 10.1109/TBME.2014.2359372] [Medline: 25252274]
47. Renevey P, Vetter R, Krauss J, Celka P, Depeursinge Y. Wrist-located Pulse Detection Using IR Signals, Activity and Nonlinear Artifact Cancellation. In: Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2001 Presented at: EMB'01; October 25-28, 2001; Istanbul, Turkey p. 3030-3033. [doi: 10.1109/iembs.2001.1017437]
48. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart rate variability analysis software. Comput Methods Programs Biomed 2014;113(1):210-220. [doi: 10.1016/j.cmpb.2013.07.024] [Medline: 24054542]
49. Tarvainen MP, Niskanen JP. Heart Rrate Variability - Kubios HRV. 2012. Kubios HRV Version 2.1 - User's Guide. URL: https://www.meditech.nu/files/2017-02/1486986930_manual-kubios-hrv-2.1.pdf [accessed 2018-12-01]

50. Tognetti S, Cenci I, Resnati D, Garbarino M, Lai M. United States Patent. 2017. United States Patent - US 9,833,155. URL: https://patentimages.storage.googleapis.com/00/1f/7f/dd5e9f97a8f8af/US9833155.pdf [accessed 2019-01-09]
51. Kim KK, Lim YG, Kim JS, Park KS. Effect of missing RR-interval data on heart rate variability analysis in the time domain. Physiol Meas 2007 Dec;28(12):1485-1494. [doi: 10.1088/0967-3334/28/12/003] [Medline: 18057513]
52. Kim KK, Kim JS, Lim YG, Park KS. The effect of missing RR-interval data on heart rate variability analysis in the frequency domain. Physiol Meas 2009 Oct;30(10):1039-1050. [doi: 10.1088/0967-3334/30/10/005] [Medline: 19713596]
53. Jovic A, Bogunovic N. Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. Artif Intell Med 2011 Mar;51(3):175-186. [doi: 10.1016/j.artmed.2010.09.005] [Medline: 20980134]
54. Antelmi I, de Paula RS, Shinzato AR, Peres CA, Mansur AJ, Grupi CJ. Influence of age, gender, body mass index, and functional capacity on heart rate variability in a cohort of subjects without heart disease. Am J Cardiol 2004 Feb 1;93(3):381-385. [doi: 10.1016/j.amjcard.2003.09.065] [Medline: 14759400]
55. Searle SR. Linear Models. First Edition. New York: Wiley-interscience; 1997.
56. Tang W, He HT, Tu XM. Applied Categorical and Count Data Analysis. New York: Chapman And Hall/CRC; 2012.
57. Wang H, Peng J, Wang B, Lu X, Zheng JZ, Wang K, et al. Inconsistency between univariate and multiple logistic regressions. Shanghai Arch Psychiatry 2017 Apr 25;29(2):124-128 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.217031] [Medline: 28765686]
58. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, 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 6;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
59. Kasper JD, Chan KS, Freedman VA. Measuring physical capacity: an assessment of a composite measure using self-report and items. J Aging Health 2017 Mar;29(2):289-309 [FREE Full text] [doi: 10.1177/0898264316635566] [Medline: 26965083]
60. Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH. Discrepancies between self-reported and observed physical function in the elderly: the influence of response shift and other factors. Soc Sci Med 1999 Jun;48(11):1549-1561. [doi: 10.1016/s0277-9536(99)00048-9] [Medline: 10400256]
61. Wulsin LR, Horn PS, Perry JL, Massaro JM, D'Agostino RB. Autonomic imbalance as a predictor of metabolic risks, cardiovascular disease, diabetes, and mortality. J Clin Endocrinol Metab 2015 Jun;100(6):2443-2448. [doi: 10.1210/jc.2015-1748] [Medline: 26047073]
62. Thayer JF, Ahs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012 Feb;36(2):747-756. [doi: 10.1016/j.neubiorev.2011.11.009] [Medline: 22178086]
63. Chandra P, Sands RL, Gillespie BW, Levin NW, Kotanko P, Kiser M, et al. Predictors of heart rate variability and its prognostic significance in chronic kidney disease. Nephrol Dial Transplant 2012 Feb;27(2):700-709 [FREE Full text] [doi: 10.1093/ndt/gfr340] [Medline: 21765187]
64. Farrell TG, Bashir Y, Cripps T, Malik M, Poloniecki J, Bennett ED, et al. Risk stratification for arrhythmic events in postinfarction patients based on heart rate variability, ambulatory electrocardiographic variables and the signal-averaged electrocardiogram. J Am Coll Cardiol 1991 Sep;18(3):687-697 [FREE Full text] [doi: 10.1016/0735-1097(91)90791-7] [Medline: 1822090]
65. Kang HG, Mahoney DF, Hoenig H, Hirth VA, Bonato P, Hajjar I, Center for Integration of Medicine and Innovative Technology Working Group on Advanced Approaches to Physiologic Monitoring for the Aged. In situ monitoring of health in older adults: technologies and issues. J Am Geriatr Soc 2010 Aug;58(8):1579-1586. [doi: 10.1111/j.1532-5415.2010.02959.x] [Medline: 20646105]
66. Singh N, Moneghetti KJ, Christle JW, Hadley D, Plews D, Froelicher V. Heart rate variability: an old metric with new meaning in the era of using mhealth technologies for health and exercise training guidance. Part one: physiology and methods. Arrhythm Electrophysiol Rev 2018 Aug;7(3):193-198 [FREE Full text] [doi: 10.15420/aer.2018.27.2] [Medline: 30416733]

Abbreviations

BMI: body mass index
CCHSC: continuing care senior housing community
CIRS-G: Cumulative Illness Rating Scale for Geriatrics
ECG: electrocardiogram
GEE: generalized estimating equation
HRPP: Human Research Protections Program
HRV: heart rate variability
PPG: photoplethysmography
SF-36: Short Form 36
SPPB: Short Physical Performance Battery
TI: triangular index

TUG: Timed Up and Go
VIF: variance inflation factor

Edited by G Eysenbach; submitted 19.02.19; peer-reviewed by L Rutter, A Kamisalic; comments to author 13.05.19; revised version received 16.07.19; accepted 19.08.19; published 23.10.19.

Please cite as:
Graham SA, Jeste DV, Lee EE, Wu TC, Tu X, Kim HC, Depp CA
Associations Between Heart Rate Variability Measured With a Wrist-Worn Sensor and Older Adults' Physical Function: Observational Study
JMIR Mhealth Uhealth 2019;7(10):e13757
URL: http://mhealth.jmir.org/2019/10/e13757/
doi:10.2196/13757
PMID:31647469

Corrigenda and Addenda

Metadata Correction: A Mobile Health Wallet for Pregnancy-Related Health Care in Madagascar: Mixed-Methods Study on Opportunities and Challenges

Nadine Muller[1,2], MD; Peter Martin Ferdinand Emmrich[3], PhD; Elsa Niritiana Rajemison[1], MEc; Jan-Walter De Neve[1], MD, ScD; Till Bärnighausen[1,4,5], MD, ScD; Samuel Knauss[6*], MD; Julius Valentin Emmrich[6*], MD, MPhil

[1]Heidelberg Institute of Global Health, Medical Faculty and University Hospital, University of Heidelberg, Heidelberg, Germany

[2]Department of Infectious Diseases and Pulmonary Medicine, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[3]Biosciences Eastern and Central Africa Hub, International Livestock Research Institute, Nairobi, Kenya

[4]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[5]Africa Health Research Institute, Mtubatuba, KwaZulu-Natal, South Africa

[6]Department of Experimental Neurology and Center for Stroke Research, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Nadine Muller, MD
Department of Infectious Diseases and Pulmonary Medicine
Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30450665296
Email: nadine.muller@charite.de

Related Article:

Correction of: https://mhealth.jmir.org/2019/3/e11420

(JMIR Mhealth Uhealth 2019;7(10):e16190) doi:10.2196/16190

In "A Mobile Health Wallet for Pregnancy-Related Health Care in Madagascar: Mixed-Methods Study on Opportunities and Challenges" (JMIR Mhealth Uhealth 2019;7(3):e11420), authors Knauss and Emmrich should have been denoted as having contributed equally to the paper but were not. A footnote clarifying equal contribution has now been added to each of the aforementioned authors.

The correction will appear in the online version of the paper on the JMIR website on October 14, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 09.09.19; this is a non–peer-reviewed article;accepted 18.09.19; published 14.10.19.

Please cite as:
Muller N, Emmrich PMF, Rajemison EN, De Neve JW, Bärnighausen T, Knauss S, Emmrich JV
Metadata Correction: A Mobile Health Wallet for Pregnancy-Related Health Care in Madagascar: Mixed-Methods Study on Opportunities and Challenges
JMIR Mhealth Uhealth 2019;7(10):e16190
URL: https://mhealth.jmir.org/2019/10/e16190
doi:10.2196/16190
PMID:31613791

Original Paper

Depression Prediction by Using Ecological Momentary Assessment, Actiwatch Data, and Machine Learning: Observational Study on Older Adults Living Alone

Heejung Kim[1,2], GNP-BC, RN, PhD; SungHee Lee[3], PhD; SangEun Lee[4], PhD; Soyun Hong[1], RN, MSN; HeeJae Kang[3], BS; Namhee Kim[1], RN, MPH

[1]College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]Mo-Im Kim Nursing Research Institute, Yonsei University, Seoul, Republic of Korea

[3]BRFrame Inc, Seoul, Republic of Korea

[4]Health-IT Acceleration Platform Technology Innovation Center, College of Medicine, Yonsei University Health System, Seoul, Republic of Korea

Corresponding Author:
Heejung Kim, GNP-BC, RN, PhD
College of Nursing
Yonsei University
Room #603
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3273
Fax: 82 2 2227 8303
Email: hkim80@yuhs.ac

Abstract

Background: Although geriatric depression is prevalent, diagnosis using self-reporting instruments has limitations when measuring the depressed mood of older adults in a community setting. Ecological momentary assessment (EMA) by using wearable devices could be used to collect data to classify older adults into depression groups.

Objective: The objective of this study was to develop a machine learning algorithm to predict the classification of depression groups among older adults living alone. We focused on utilizing diverse data collected through a survey, an Actiwatch, and an EMA report related to depression.

Methods: The prediction model using machine learning was developed in 4 steps: (1) data collection, (2) data processing and representation, (3) data modeling (feature engineering and selection), and (4) training and validation to test the prediction model. Older adults (N=47), living alone in community settings, completed an EMA to report depressed moods 4 times a day for 2 weeks between May 2017 and January 2018. Participants wore an Actiwatch that measured their activity and ambient light exposure every 30 seconds for 2 weeks. At baseline and the end of the 2-week observation, depressive symptoms were assessed using the Korean versions of the Short Geriatric Depression Scale (SGDS-K) and the Hamilton Depression Rating Scale (K-HDRS). Conventional classification based on binary logistic regression was built and compared with 4 machine learning models (the logit, decision tree, boosted trees, and random forest models).

Results: On the basis of the SGDS-K and K-HDRS, 38% (18/47) of the participants were classified into the probable depression group. They reported significantly lower scores of normal mood and physical activity and higher levels of white and red, green, and blue (RGB) light exposures at different degrees of various 4-hour time frames (all $P<.05$). Sleep efficiency was chosen for modeling through feature selection. Comparing diverse combinations of the selected variables, daily mean EMA score, daily mean activity level, white and RGB light at 4:00 pm to 8:00 pm exposure, and daily sleep efficiency were selected for modeling. Conventional classification based on binary logistic regression had a good model fit (accuracy: 0.705; precision: 0.770; specificity: 0.859; and area under receiver operating characteristic curve or AUC: 0.754). Among the 4 machine learning models, the logit model had the best fit compared with the others (accuracy: 0.910; precision: 0.929; specificity: 0.940; and AUC: 0.960).

Conclusions: This study provides preliminary evidence for developing a machine learning program to predict the classification of depression groups in older adults living alone. Clinicians should consider using this method to identify underdiagnosed

subgroups and monitor daily progression regarding treatment or therapeutic intervention in the community setting. Furthermore, more efforts are needed for researchers and clinicians to diversify data collection methods by using a survey, EMA, and a sensor.

(JMIR Mhealth Uhealth 2019;7(10):e14149) doi:10.2196/14149

KEYWORDS

elderly; one-person household; depression; ecological momentary assessment; actigraphy; machine learning

Introduction

Challenges of Geriatric Depression

Depression is one of the most prevalent mental health problems in older adults. Globally, it was the third leading cause of years lived with disability in 2015, increasing by 18.4% between 2005 and 2015 [1]. In particular, older adults living alone are most vulnerable to depression compared with those living with others [2,3] because 30.2% of Koreans living alone are reported to have significant symptoms of depression [4]. In Korea, the economic burden of individuals with depression increased by about 27.1% from 2009 to 2013, with an estimated total cost of Korean $27.1 billion [5]. Thus, early and accurate diagnosis is critical to initiate timely treatment and reduce further disease burden; however, it is difficult to diagnose geriatric depression [1,2].

Underdiagnosis or inaccurate diagnosis of depression in older adults living alone remains a challenge [2]. Self-report and clinical interview are the primary methods to diagnose depression [6]; however, there are several challenges to overcome when diagnosing older adults living alone. First, no proxy currently exists that uses objective observations for reporting older adults' depressive symptoms in their daily lives. Second, existing self-reporting instruments have a limitation to assess mood variability in response to daily stress in the natural environment, even the validity and reliability [6]. Third, atypical symptoms of depression are more common in older adults compared with younger adults [2,6,7]. Finally, older Asian adults are often hesitant to report their depressed mood and depressive symptoms because of social stigma or misconceptions [8]. Thus, it is vital to use diverse methods to collect additional data to fill the gaps contributing to the underdiagnosis of geriatric depression [2,6].

Potential Methods to Identify Geriatric Depression

Ecological momentary assessment (EMA) has been suggested as a promising instrument because it can detect an individual's real-time experiences and mood in real-world settings over time and in different situations [9]. Currently, clinicians rely on the retrospective reports of patients' depressed mood in unfamiliar examination rooms using standardized instruments [6,10]. However, EMA allows individuals to report their momentary mood in the *right now and here*, multiple times in their real and familiar environment [9,11], rather than in artificial settings or laboratories. Thus, health care professionals can collect high ecologically valid data that are more specific to the older adult's contextual situation and readily applicable to lifestyle modification [9,11,12].

As some older adults have difficulty using high-tech devices, one of the challenges with using EMA by older adults is the limited choices available when choosing a device [9]. Although a diverse number of smart devices are available, a wrist-worn Actiwatch is one method for collecting ecological momentary data [11] and other types of data. It has been primarily used to measure objective sleep and motor activity [13,14]. The usefulness of actigraphy has been well recognized and validated in individuals with depression [15] because of its familiarity, portability, and simple operation that involves pressing the device button to input the data [13]. Although actigraph has been suggested as being an advantageous device for use with older adults [16], few studies have used actigraphy to measure momentary mood or related problems in older adults with depression [17,18].

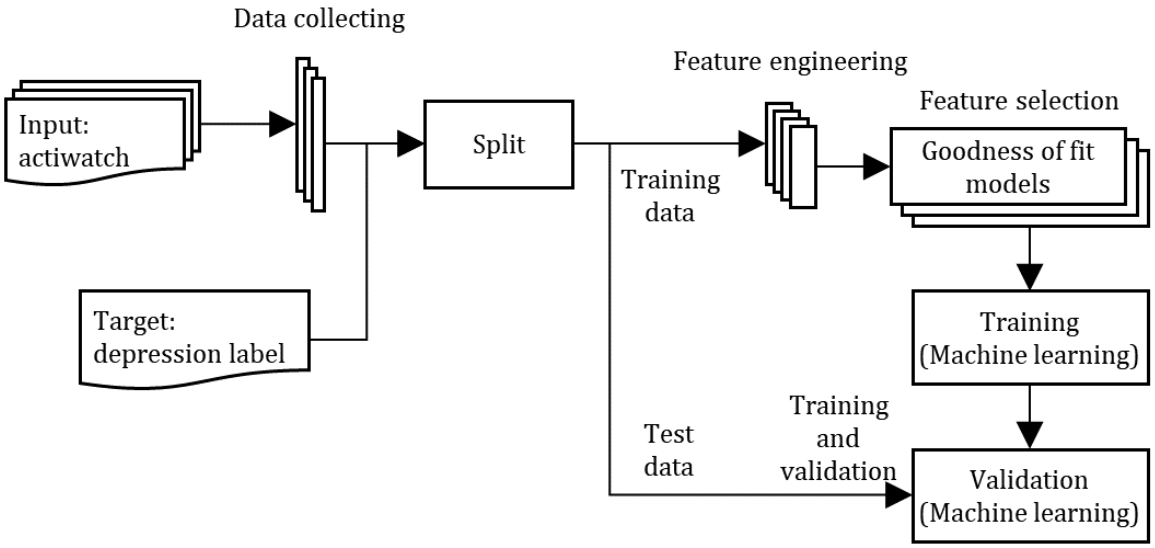
EMA and actigraphy data require a new analytic approach rather than conventional analysis. Numerous studies have used machine learning with wearable sensor data to improve symptom detection and for monitoring in diverse populations of neuropsychiatric patients living in real-time and in diverse life contexts [19-21]. Specifically, the assessment of suicide risk and emotional distress was identified as a possibly applicable area for using social media text mining [22], predicting negative emotions using mobile phone usage pattern [23], and detecting everyday behavior related to clinically meaningful levels of depression using sensor data [24]. Machine learning involves an objective, data-driven, and situation-independent analysis, and it has also been applied in the analysis of sensor data for automated assessment of mental health [24-26]. Machine learning analytic techniques that computationally mine meaning from data and classify, detect, and segment meaningful patterns, associations, relationships, and trends between variables and simulation are used to build predictive and optimization models [25,27]. These data analytics can equally be applied to small data to extract and model insights [25,27]. Therefore, depression is an ideal construct to establish, validate, and clinically apply machine learning approaches using sensor data. Thus, this study aimed to develop a supervised machine learning algorithm to predict the classification of depression groups among older adults living alone using EMA data from older adults living with a depressed mood.

Methods

Overview of Study Design

On the basis of expert opinion [28] and a methodological guide [27], this study was conducted in 4 steps: (1) collecting data associated with depression, (2) data processing and representation, (3) data modeling (ie, feature engineering and selection), and (4) training and validation of the prediction model (Figure 1).

Figure 1. Construction prediction model.



Data Collection

Sample

We recruited a convenience sample of 56 older adults living alone via a community health center in Korea between May 2017 and January 2018. Inclusion criteria were as follows: the participants (1) were aged 65 to 94 years, (2) understood Korean, (3) lived as a single household in the community, and (4) had at least mild levels of depressive symptoms (score ≥5 on the Short Geriatric Depression Scale). We excluded 3 individuals with significant cognitive impairment and a high risk of suicide according to the Korean versions of the Mini-Mental Status Examination [29] and Crisis Triage Ration Scale [30]. In addition, the data from 6 individuals were excluded from the analysis because of refusal to wear or incomplete wearing of an Actiwatch (n=5) and data loss because of device error (n=1). Thus, the final sample comprised 47 participants who constantly wore an Actiwatch and reported sufficient numbers of EMA (ie, at least once a day) for 2 weeks [31]. All participants provided written informed consent, and the institutional review board of the affiliated university (IRB 2017-0007-1) approved the study.

Depression Measures

To classify the depression and nondepression groups, the Korean versions of the Short Geriatric Depression Scale (SGDS-K) and the Hamilton Depression Rating Scale (K-HDRS) were used to assess self-reported and clinician-observed symptoms of depression at baseline and 2 weeks. The SGDS-K is a 15-item instrument to assess subjective depressive symptoms (0=symptom absent and 1=symptom present). The total score ranges from 0 to 15; a higher score indicates more severe levels of depression [32]. The K-HDRS is a 17-item instrument that is the most widely used observer rating scale for observing depressive symptoms and severity. The total score ranges from 0 to 52, and a higher score indicates more severe depression [33]. The diagnostic validity of SGDS-K and K-HDRS for screening has been reported in previous studies [33,34].

Actigraphy Data and Ecological Momentary Assessment

Physical activity and ambient light exposure were measured using a wrist-worn Actiwatch (Actiwatch Spectrum PRO, Philips Respironics). It is considered to be a valid and reliable accelerometer that continually detects wrist movements that reflect activities; sleep-wake patterns; light exposure to the red, green, and blue (RGB) spectral regions; and broad-spectrum white light [35,36]. Data were collected continuously in 30-second epochs for 14 consecutive days. Participants wore the Actiwatch on the nondominant wrist all the time and were instructed to take it off only when taking a bath or for a few minutes as needed. Furthermore, participants were instructed that long sleeves should not cover the light sensor in the Actiwatch. We used Actiware software (Philips Respironics) to export the data.

Participants were instructed to rate their momentary mood using a button on the Actiwatch for 2 weeks. When the participants pressed the button, an increasing number was shown in the front window of the Actiwatch. The mood level was measured on a 10-point Likert scale (ranging from 1=very depressed to 10=not depressed) on the individualized preset time, as EMA reporting time in a day should be set in consideration of each individual's lifestyle and convenience [37]. The Actiwatch included options for audible and vibrating alarms that were set to appear 4 times a day as a means to remind participants to complete the EMA regarding their current depressed mood. When a participant missed the alarm to complete the EMA, they received additional alarms over 5-min intervals until they responded. Research staff also provided a follow-up phone call to all participants after 1 week, and additional contact was provided for troubleshooting on an *as-needed* basis. Participants received gifts that were worth US $10 for completing the 2-week data collection.

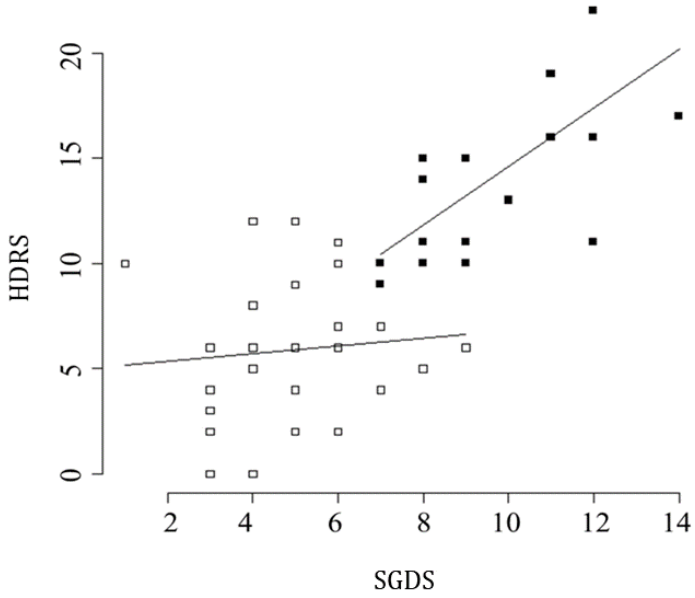
Data Processing and Representation

Defining Depression

The diagnosis of depression in older adults living alone was made using a decision-making model that is primarily used in

the machine learning method [38]. To assess depression, SGDS-K and K-HDRS were administered at baseline and 2 weeks later. The data for a probable diagnosis of depression were based on the second measure at 2 weeks, during which the correlation between SGDS-K and K-HDRS was the highest (r=0.753; *P*=.01). To develop a diagnostic model, participants were classified into 2 groups: depression (SGDS-K≥7 and K-HDRS≥8) and nondepression (SGDS-K≤6 and K-HDRS≤7) groups [39-41]. Finally, 18 older adults were classified into the depression group and 29 were classified into the nondepression group (Figure 2).

Figure 2. Correlations between Short Geriatric Depression Scale (SGDS) and Hamilton Depression Rating Scale (HDRS) scores by the depression groups. White squares mean nondepression group, whereas the black squares mean depression group.



Data Processing

Data were collected through the Actiwatch for 2 weeks. Data preprocessing began with the downloading of raw data and conversion into usable data. In a participant's activity time series, continuous data were collected by the Actiwatch every 30 seconds. Triaxial data were calculated as an activity count. On the basis of the characteristics of signal data collected in the time series at 30-second intervals, null or abnormal values because of user or device error were checked and excluded from the analysis.

If a participant removed the device, the triaxial accelerometer would record values of 0 to indicate the duration of time for which it was not worn. Natural human behavior involves micromovements that are sensed by the accelerometer even during sleep, and therefore, periods with a continuous absence of movement indicate device removal. The raw accelerometer data were aggregated into minute-by-minute epochs using a script written in R (version 3.3.1). Data were considered as missing or being excluded from the analysis when (1) there were 0 measures in 5 min despite the indication of the *wearing* status or (2) more than 60% of a participant's data were missing in a day.

The processing of EMA scores was performed as follows: (1) exploring the pattern of each participant's EMA scores of depressed mood at 4 different time points each day, (2) comparing the 4 EMA scores and the grand mean scores between 2 groups, (3) examining the overall patterns between 2 groups, and (4) computing a composite score of daily EMA score.

Data Modeling of Depression Prediction

Feature Engineering

Feature engineering is a step of parameterizing the collected raw data. It is a data mining process that quantitatively and qualitatively characterizes the statistics of the measured values for each participant through data aggregation or pattern analysis [38]. Moreover, it is important to determine additional variables to consider for the prediction model during feature engineering [38]. The variables to be included in the prediction model are the original variables included in the existing data and the new types of variables created using these primitive variables. Through the feature engineering process, the existing and new variables are identified to specify the characteristics of the participants and improve the accuracy of the prediction models [38].

After examining the daily fluctuations in the EMA, activity, and ambient light exposure, we compared both groups' 4-hour mean differences in the selected variables using a Mann-Whitney U test and time series plot. To find additional variables to consider for the prediction model, we tested diverse types of sleep parameters, such as total time in bed, total sleep time, sleep efficiency, and wake-after-sleep-onset (WASO).

Feature Selection

This process selects and filters significant variables used in the prediction model. Owing to the small sample size, unnecessary

variables in the process of predicting depression may decrease the model's degree of freedom, which can negatively affect the explanatory power or the normal operation of the model [38]. Therefore, it is necessary to evaluate the significance of the processed variables through hypothesis testing and logistic regression analysis to select the most efficient combination of variables [38].

The actual selection of the variables was performed by a mean difference test of the statistics of the depression and nondepression groups and by logistic regression analysis based on various combinations of variables to examine the fitness of the models and determine the final combination of explanatory variables to be used in the prediction [38]. We selected EMA score, activity, and ambient light exposure and sleep efficiency to perform a binary logistic regression.

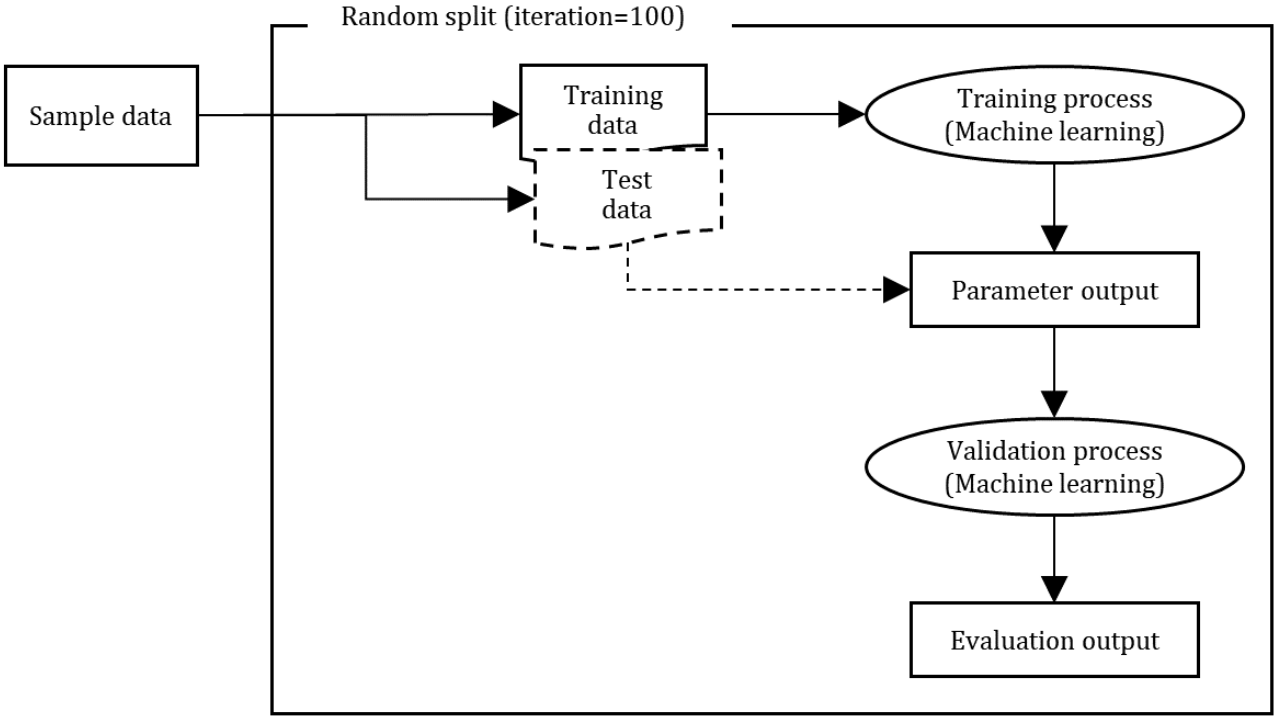
Training and Validation Test

We evaluated training and validation data of the predictive model to classify depression groups based on the previously selected explanatory variables through the cross-validation process and assessed the validity of the final model [38]. Data from 47 participants were divided into training data and test data, and the machine learning method was applied to calculate the predictive power of the model. Specifically, the parameter value of the prediction model was calculated from the learning data and was applied to the test data, and the prediction level was evaluated using rxLogit function in Microsoft R package.

Data Partitioning

To train our models without overfitting and test their subsequent performance, we randomly partitioned the dataset. Each time series was randomly assigned to a partition while maintaining an even class distribution of the target variable, depression group. Data were split at a 0.65:0.35 ratio for training and testing, respectively. Thus, the data were divided as follows: 30 as training data (n=20 for the nondepression group and n=10 for the depression group) and 17 as test data (n=9 for the nondepression group and n=8 for the depression group) in consideration of the sample size. Ideally, to ensure the cross-validation of the model and the validity of the evaluation, the data resplitting process should be repeated 100 times or more at random (Figure 3).

Figure 3. Machine learning—training, validation, and test.



Training the Models

Machine learning techniques that were used for training and validation included the logit, decision tree, boosted trees, and random forest models. To test the validity of the prediction model, several indices were used such as accuracy, precision, recall (or sensitivity), specificity, F score, and area under receiver operating characteristic curve (AUC) in comparison between the machine learning model and a logistic regression, which are mainly used in the classification model [38]. Each indicator was calculated in the Confusion matrix. Our model showed that depression was positive and nondepression was negative.

Results

Descriptive Statistics of the Sample

The average age of the 47 participants was 78 (SD 5.24) years; they were mainly women (44/47, 94%), had less than high school education (42/47, 89%), and had a moderate socioeconomic status (35/47, 74%). The sample characteristics of the 6 participants who were excluded from the analyses did

not significantly differ from those of the 47 participants included in the final analyses.

On the basis of the traditional depression assessment tools (SGDS-K and K-HDRS), 38% (18/47) of the participants were classified into the depression group. We performed the Mann-Whitney U test to compare the 2-week grand means of the depression and nondepression groups, and significant differences were identified. The depression group reported a significantly lower score of nondepressed mood and physical activity and higher levels of exposure to white and RGB light (all *P* values <.01). Data were used to identify the pattern of time series of signal data and for reference for interpretation (Table 1).

Table 1. Mean differences tested by the Mann-Whitney U test.

Characteristics	Nondepression group (n=29), mean (SD)	Depression group (n=18), mean (SD)	Differences, mean (SD)	*P* value
Ecological momentary assessment	6.6 (1.55)	5.1 (1.64)	−1.5 (1.58)	.004
Activity (counts)	90.5 (32.76)	67.4 (18.82)	−23.1 (28.31)	.003
Light exposure (lux)				
White	54.0 (36.54)	81.5 (39.98)	27.5 (37.88)	.008
Red	69.9 (54.12)	103.6 (63.95)	33.7 (58.03)	.03
Green	60.1 (43.09)	99.9 (55.35)	39.8 (48.09)	.005
Blue	38.9 (28.84)	62.0 (34.33)	23.1 (31.03)	.006

Feature Engineering and Selection

Mean Differences in Different Sections of Time

To obtain a descriptive picture of the daily fluctuations in EMA, activity, and ambient light exposure and compare the 4-hour mean differences of the 2 groups, we performed the Mann-Whitney U test and time series plot. The depression group reported lower levels of EMA scores and daily activity throughout the day. However, there were higher levels of white and RGB light exposures in the depression group compared with the nondepression group.

On further examination of different time frames, both groups showed the highest levels of activity in the morning between 8:00 am and 12:00 pm. However, the highest levels of white and RGB light exposures were observed between 12:00 pm and 4:00 pm in the nondepression group and 8:00 am and 12:00 pm in the depression group. More activities were observed for the depression group after 9:00 pm compared with the nondepression group, although the difference was not significant (Figure 4). This finding suggests that 4-hour data could be considered for modeling (Table 2).

Figure 4. Daily time series plot of depression versus nondepression. Activity patterns for depression and nondepression groups. Y axis is average raw activity counts depending on each time of day and X axis is the time of day.

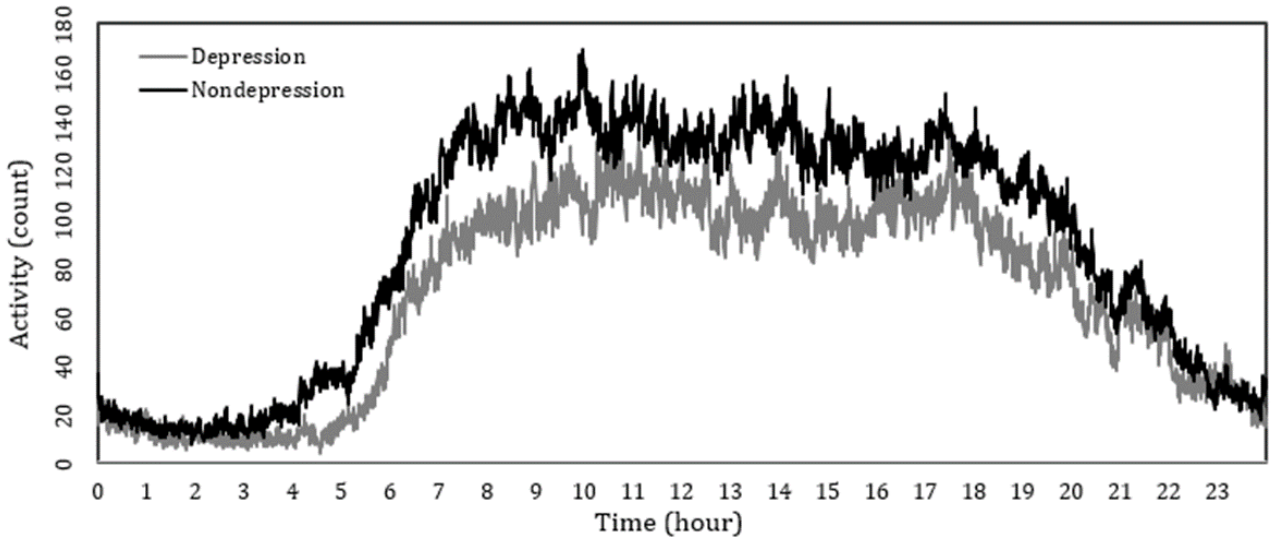


Table 2. Mean differences in different sections of time.

Variables	12:00 am-4:00 am	4:00 am-8:00 am	8:00 am-12:00 pm	12:00 pm-4:00 pm	4:00 pm-8:00 pm	8:00 pm-12:00 am
Ecological momentary assessment						
Nondepression	—[a]	6.4	6.6	6.6	6.6	—
Depression	—	5.0	5.2	5.2	5.1	—
Difference	—	−1.4	−1.4	−1.4	−1.5	—
P value	—	.02	.009	.01	.005	—
Activity						
Nondepression	16.4	78.5	141.2	133.0	120.1	54.1
Depression	11.1	48.1	107.1	99.4	94.8	44.1
Difference	−5.3	−30.4	−34.1	−33.6	−25.3	−10.0
P value	.36	.02	.01	.005	.06	.20
White light						
Nondepression	1.0	14.5	110.3	137.4	50.7	10.4
Depression	0.7	18.1	196.3	161.6	91.7	20.6
Difference	−0.3	3.6	86.0	24.2	41.0	10.2
P value	.51	.42	.02	.08	.005	.10
Red light						
Nondepression	0.6	17.1	150.0	181.7	64.2	5.8
Depression	0.6	20.9	269.6	204.9	112.1	13.4
Difference	0.0	3.8	119.6	23.2	47.9	7.6
P value	.79	.43	.049	.12	.02	.07
Green light						
Nondepression	1.1	17.0	125.0	154.2	54.3	8.9
Depression	1.4	21.4	250.2	194.7	103.0	28.5
Difference	0.3	4.4	125.2	40.5	48.7	19.6
P value	.97	.31	.01	.05	.003	.09
Blue light						
Nondepression	0.4	10.3	81.9	101.5	34.7	4.7
Depression	0.4	13.7	154.6	126.2	64.8	12.3
Difference	0.0	3.4	72.7	24.7	30.1	7.6
P value	.77	.33	.01	.07	.006	.08

[a]Not applicable.

Additional Variable: Sleep Efficiency

As there were no differences in activity levels at night (see Figure 4) and no significant difference in other sleep components, sleep efficiency was chosen for modeling. Sleep efficiency is defined as the ratio of the total sleep time at night compared with the total amount of time spent in bed, which reflects night-time activity and the quality of sleep. We calculated sleep efficiency based on sleep time and WASO based on sleep component equations (Multimedia Appendix 1) [42].

In this study, activity for more than 5 consecutive minutes during the sleep time was considered as *wakefulness*. The total sum of all moments of wakefulness is referred to as WASO, when smoothing intermittent movements of night-time activity. A cutoff of 85% or higher was used to determine good sleep efficiency [42]. On the basis the mean differences in sleep efficiency of the 2 groups, the nondepression group had a sleep efficiency of 86% (25/29), which was higher than the 85% cutoff, whereas the depression group showed 83% (15/18), lower than the cutoff. The between group difference was different at the *P*=.08 level regarding sleep efficiency.

A binary logistic regression was performed to explore the best model fit using various combinations of variables (EMA score, activity, white and RGB light exposure, and sleep efficiency) depending on depression groups (nondepression group=0 and

depression group=1). We compared diverse models using different means based on EMA and activity data. For example, we compared daily, morning-evening means, and all 4 scores as well as 4 different times of light exposure and other sleep components. Finally, we reached the conclusion that the variables of the final model were the best when using daily mean EMA score, daily mean activity level, white and RGB light at 4:00 pm to 8:00 pm exposure, and daily sleep efficiency. To consider collinearity between white and RGB light change and the small sample size, a computing score was calculated integrating means of white and RGB light between 4:00 pm and 8:00 pm. Table 3 shows goodness of fit depending on each variable. Our findings suggest that lower activity, higher EMA depressed mood, and exposure to white and RGB were associated with a higher likelihood of being classified as a depression group. Furthermore, this model showed a good model fit (accuracy: 0.705; precision: 0.770; specificity: 0.859; and AUC: 0.754) except 1 false-positive case.

Table 3. Estimates of logistic regression.

Variables, estimator	Estimates	*P* value
Intercept		.02
Coefficient	35.200	
Activity		.02
Coefficient	−0.095	
Odds ratio (95% CI)	0.910 (0.84-0.98)	
Marginal effect	−0.016	
Computing score of white and red, green, and blue light		.01
Coefficient	0.024	
Odds ratio (95% CI)	1.025 (1.01-1.04)	
Marginal effect	0.004	
Sleep efficiency		.08
Coefficient	−25.300	
Odds ratio (95% CI)	<0.001 (<0.001-14.41)	
Marginal effect	−4.362	
Ecological momentary assessment		.01
Coefficient	−2.299	
Odds ratio (95% CI)	0.100 (0.02-0.58)	
Marginal effect	−0.397	

Comparison of Four Machine Learning Methods

Table 4 shows the validation and test results for the prediction model based on machine learning. All variables included in the logistic regression were utilized (EMA score, activity, white and RGB light exposure, and sleep efficiency) to classify depression groups (nondepression group=0 and depression group=1). For cross-validation, bootstrapping with data partitioning and training models was performed 100 times. Among the 4 methods, the logistic regression model (accuracy: 0.910; precision: 0.929; and specificity: 0.940) had the best fit compared with the boosted trees and random forest models. Decision tree suffers from possible overestimated model fit because of sample size. boosted trees model showed relatively better fit compared with decision tree because of the modification process to correct errors through pruning. Random forest seemed less suitable for this study because of the small sample size and the limited number of tested variables [38].

Table 4. Evaluation metrics.

Evaluation metrics	Logistic regression		Decision tree		Boosted trees		Random forest	
	Mean (SD)	Minimum-maximum	Mean (SD)	Minimum-maximum	Mean (SD)	Minimum-maximum	Mean (SD)	Minimum-maximum
Accuracy	0.91 (0.007)	0.87-0.92	0.72 (0.03)	0.64-0.78	0.78 (0.04)	0.68-0.87	0.67 (0.05)	0.53-0.75
Precision	0.93 (0.01)	0.88-0.95	0.74 (0.04)	0.63-0.81	0.84 (0.06)	0.69-0.95	0.72 (0.11)	0.50-1.02
Recall/sensitivity	0.88 (0.01)	0.84-0.90	0.63 (0.05)	0.55-0.73	0.66 (0.06)	0.57-0.68	0.39 (0.07)	0.23-0.50
Specificity	0.94 (0.01)	0.90-0.96	0.80 (0.04)	0.72-0.88	0.84 (0.05)	0.78-0.98	0.88 (0.06)	0.80-1.04
F score	0.90 (0.01)	0.86-0.92	0.68 (0.03)	0.59-0.75	0.74 (0.05)	0.62-0.83	0.51 (0.05)	0.38-0.62
Area under receiver operating characteristic curve	0.96 (0.03)	0.91-0.99	0.66 (0.03)	0.61-0.71	0.84 (0.03)	0.79-0.89	0.71 (0.03)	0.69-0.79

Discussion

Principal Findings

The primary purpose of this study was to identify factors associated with geriatric depression in older adults living alone. We focused on developing a prediction model to classify the depression groups (probable depression vs nondepression) among older adults living alone. Along with the conventional instruments for depression screening, EMA of daily mood, actigraphy data of activity, and light exposure were utilized. Comparing diverse combinations of the selected variables, daily mean EMA score, daily mean activity, daily sleep efficiency, and exposure to white and RGB light between 4:00 pm and 8:00 pm for 2 weeks were selected for modeling. The cross-validation process was used to build a prediction model. Logit model showed compatible evaluation metrics compared with the traditional binary logistic regression. The use of both EMA and sensor data seems promising to develop a machine learning model for better identification of a probable depression among older adults [25].

Comparison With Previous Studies

The depression group reported higher levels of daily depressed mood than did the nondepression group. In a small-scale study [17], the major depressive disorder (MDD) group reported higher levels of diurnal symptom patterns of negative affect with great variability than did the control group. Our study found that even the daily average of depressed mood was significantly related to the classification of depression groups determined by conventional screening [2,43]. This could help overcome a clinical challenge in diagnosing a psychiatric disorder during the first medical examination or interview by unfamiliar clinicians [2]. It can also be difficult for a person with a serious mental illness to complete the self-reporting questionnaire and for clinicians to make a quick diagnosis with limited time [6,44]. The findings of this study suggest that individuals with these difficulties may complete self-reports of depressed mood daily in their home environment. A simple question, although repetitive, might be more helpful to classify the probable depression groups as opposed to the application of complex and multiple-item questionnaires.

A significantly low level of daytime activity was observed in the depression group, similar to the findings of previous research [45]. A study [15] using actigraphy reported that patients with depression displayed less daytime motor activity than did individuals without depression. Related to the EMA of depression, high levels of momentary depressed mood had a lagged effect of prolonging the current status of being at home. Those who experience an increase in the average depressed mood the previous day tend to stay at home the following day [46]. Previous studies have reported that depressed people have decreased levels of daytime activity compared with healthy controls [2,15]. In adults aged over 60 years, depression is associated with lower levels of physical activity in global measure, compared with nondepressed individuals [47], similar to our study findings. In a meta-analysis of adults with MDD [48], more time was spent in sedentary behavior and less time in vigorous physical activities. Considering the relationship between physical activity and depression, promoting daytime physical activity could prevent the risk of depression in older adults [49].

Furthermore, sleep efficiency may be a potential factor to aid in classifying depression groups. Existing studies have confirmed the association between disrupted sleep and depression in adults [50-52] because sleep disturbance is one of the common misinterpreted symptoms of geriatric depression [6]. A study of older adults living alone found that depressed participants reported sleep-related issues [53]. However, our study did not clearly indicate that sleep efficiency was associated with the depression group because the sample size had low power to detect the significance (P<.08). Comparing with the findings of previous studies [12,54], this discrepancy should be interpreted with caution because the previous studies measured sleep components once through self-report, whereas our study assessed daily sleep efficiency for 2 weeks through repeated measures using actigraphy. Older adults living alone may have more difficulty in the early detection of sleep disturbances than do those living with others because of the absence of a bedroom partner [55]. Thus, further research using both subjective and objective measures is needed to confirm the predictive value of sleep efficiency with the larger sample.

Contrary to our expectation, higher levels of ambient light exposure were observed in the depression group, compared with the nondepression group. We expected that lower levels of ambient light exposure would be observed in the depression group compared with the nondepression group that had a less sedentary lifestyle and engaged in more outdoor activities [48]. There are limited explanations for this finding contrary to our hypothesis. A small-scale observational study [56] explored the subjective perception of lighting in depressed patients' environment. When they were asked whether the lights in the surroundings seemed dimmer than usual, depressed patients answered that they perceived the environment to be dimmer than usual. Moreover, there was a significant association between depression severity and perception of dimness. Patients with severe (65%), moderate (21%), and mild (14%) depression responded that their ambient environment appeared dimmer than usual. To compensate for the perceived dim environment, they may turn on the lights at home most of the time or stay in places with more light. In addition, a depressed individual may turn on the lights to read books or do other things to compensate their awakening time at night. This behavior may affect heart rate variability, which is associated with mood regulation through ambient biofeedback [57]. Many relaxation methods or interventions use dimmed ambient lighting to soothe the patient [58], but individuals with depression may not prefer a dimmer environment that may increase the depressed mood.

Implications for Geriatric Depression Research

Our study findings support 3 major inferences: (1) diversifying data collection methods to build an accurate model of depression prediction, (2) emphasizing the need to assess depressed mood at different time points in a day, and (3) monitoring multiple symptoms continuously. First, inactivity and poor sleep have been known to be symptoms of geriatric depression; however, the assessment of these symptoms is limited to using subjective reports of inactivity, sedentary lifestyle, or night-time sleep. Our study emphasizes the need for collecting both subjective and objective data related to mood disorders and examining intraday change patterns. For example, it is difficult to apply the Pittsburgh Sleep Quality Index global score to depressed older adults. Thus, clinicians should evaluate sleep efficiency through actigraphy as an objective measure, which may prevent the misinterpretation of a suspected depression [6]. Second, time-specific assessment is required to capture the significant features of specific data. Our study findings confirmed group differences in daytime activity, light exposure in the late afternoon, and night-time sleep. Thus, we suggest the need to collect data throughout the day and identify features closely related to depression [12,45]. Third, we suggest using sensor data, such as Actiwatch or activity tracking, for monitoring purpose. For example, strategies to improve physical activity should be included in the treatment for geriatric depression [48]. Monitoring daily activity of people with mood disorders using Actiwatch may help diagnose depression depending on the activity-based interventions [15]. In addition, our study shows some potential to use environmental data, such as ambient light, in identifying depression in the community setting. Sensor devices have a great potential to continuously capture diverse mental health–related information from the environment or context [24,25] in mental health.

Implications of Electronic Health for Older Adults

EMA could be used as a diagnostic tool for depression in older adults. It has been used to assess multiple mental health indicators related to depressed mood [9]. In particular, older adults have less accurate retrospective memory, and therefore, the use of EMA for older adults can increase the accuracy of the assessment [10,59]. EMA has also been shown to have acceptability and feasibility for older adults, even those with cognitive and emotional difficulties [60]. In particular, EMA has the advantage of continuously reporting subjective symptoms according to the individual's life patterns as digital phenotyping [44,61]. Thus, it is possible to collect personalized data, monitor the individual pattern, and examine fluctuations in the depressed mood within and between a day while reducing retrospective bias [10,11].

However, a limitation of this method is that multiple and repeated reports in a day are associated with a high dropout rate [11]. When applying EMA to older adults, sufficient device training and the practice of technical details could maximize the usability and appropriateness of EMA for older adults [60]. In our study, a dropout rate of 9% (5/53) was observed; thus, preventing dropout is important to ensure that older adults with depression complete multiple reports during the day after careful training and educating about the significance of EMA [11,16,62]. Selection of valid questions to measure momentary depressed mood and standardizing the EMA protocols are also important. Therefore, geropsychiatric clinicians should prepare to select the proper device, train older adults, and prevent dropout [16].

The paper-based diary approach has been traditionally used for older adults [6,63]; however, sometimes they do not accurately report their depressed mood because of response or retrospective bias or the social stigma associated with mental health illnesses [6,10,44,59]. Actigraphy is a noninvasive technique that provides simple and scientifically accurate data on a daily basis and can be used to improve health-related behaviors [64], thus suggesting its suitability for older adults. The 9% (5/53) dropout rate in our study indicates the acceptability and applicability of the Actiwatch for older adults, similar to a previous finding [11], especially as we identified a variety of subjective and objective factors that reflected an individual's real-world environment, such as depressed mood, activity, sleep, and different light using a ubiquitous device.

Actigraphy also measures fluctuating subjective mood on a daily basis in real time and a natural environment through continuous recording [9] along with physical and environmental data [11]. An integrated approach to passive, objective, and continuous measurement has been used in psychiatry practice and research [44]. Moreover, passive actigraphy data and active EMA responses can be useful for understanding the characteristics of depression, offering insight into biological and environmental mechanisms underlying the depression, and developing individualized interventions for older adults [65]. Considering the rapid development of technologies, various types of sensor data may be used in geropsychiatric care for

screening, monitoring, and diagnosing mental health [66-68]. Therefore, we expect that new measures, such as using sensor data, will help diagnose depression quickly and improve mental health problems.

Limitations

Our study has several limitations. First, we used 2 screening instruments that relied on subjective reports to classify individuals in the 2 depression groups because these instruments had been widely used in community settings. However, a medical diagnosis based on Diagnostic and Statistical Manual of Mental Disorders–5 may be needed to define the clinically diagnosed depression group for data processing and representation [6]. In addition, further study may replicate our model including nondepressed older adults to establish screening models in general population. Although the EMA provides mood reports of older adults at the 4 time points every day for 2 weeks, our predictive model used daily grand mean that was relatively less frequent for EMA data collection. Replication studies that utilize data variability within a day should be tested over a longer time frame and use a more convenient mode to complete EMA to reduce subjective burden. Generalization of the EMA data and levels of activity is limited because every individual has different levels of normal with reference to a depressed mood or inactivity [6,11]. Thus, EMA with a real-life tagging system will be helpful to examine to what extent the data fluctuate from the average level for each individual. In addition, studies with larger samples are needed to ensure the generalizability of our findings considering the possible use of machine learning, which is an innovative method of feature extraction from data [21]. Specifically, the limited generalizability is expected for depressed men in older populations because the majority of older adults were women in Korea. Thus, it is necessary to oversample the groups of men for the next study on this topic [4].

Conclusions

This study provides evidences to support the feasibility of machine learning in classifying depression groups based on EMA and actigraphy data. Specifically, our machine learning approach is applicable for females and elderlies with depressive mood when living alone in the community setting. The study findings provide 2 major inferences: (1) diversifying data collection methods to build an accurate model of prediction and (2) emphasizing the need for assessments at different time points in a day to obtain diverse data. Future researchers and clinicians should consider EMA and actigraphy to obtain data on daily mood, levels of activity and ambient light exposure, and sleep components in depressed older adults living alone.

Acknowledgments

The authors express their appreciation to all study participants for their time and effort in completing the surveys. In addition, they appreciate Yeonsu-gu public health centers (Incheon, South Korea) for their support in recruiting participants. This study was supported by Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education (NRF-2016R1D1A1B03932013) and the Ministry of Trade, Industry, and Energy Grant funded by the Korean government (Medical Device R&D Platform Project, Project Number: 10060085).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Equation to calculate sleep efficiency based on sleep time and wake-after-sleep-onset.
[PDF File (Adobe PDF File), 26 KB - mhealth_v7i10e14149_app1.pdf]

References

1. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 8;388(10053):1545-1602. [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
2. Sözeri-Varma G. Depression in the elderly: clinical features and risk factors. Aging Dis 2012 Dec;3(6):465-471. [FREE Full text] [Medline: 23251852]
3. Kim H, Kwon S, Hong S, Lee S. Health behaviors influencing depressive symptoms in older Koreans living alone: secondary data analysis of the 2014 Korean longitudinal study of aging. BMC Geriatr 2018 Aug 20;18(1):186. [FREE Full text] [doi: 10.1186/s12877-018-0882-4] [Medline: 30126354]
4. Korea Institute for Health and Social Affairs. 2017. 2017 Korean Elderly Survey. URL: https://data.kihasa.re.kr/micro/subject_view.jsp?grp_seq=64&project_seq=673 [accessed 2019-02-28] [WebCite Cache ID 771z8Yvr6]
5. Mobile Health Insurance Review & Assessment Service. 2017. Disease Information by Statistics. URL: http://www.hira.or.kr/dummy.do?pgmid=HIRAA030502030000&ilnsCtgyNm=sixtyAge&sortSno=46 [accessed 2019-02-28] [WebCite Cache ID 76rSB7ztS]
6. Balsamo M, Cataldi F, Carlucci L, Padulo C, Fairfield B. Assessment of late-life depression via self-report measures: a review. Clin Interv Aging 2018;13:2021-2044. [FREE Full text] [doi: 10.2147/CIA.S178943] [Medline: 30410319]

7. World Health Organization. 2017. Mental Health of Older Adults. URL: https://www.who.int/en/news-room/fact-sheets/detail/mental-health-of-older-adults [accessed 2019-03-14] [WebCite Cache ID 76r9M7qrr]
8. Jang Y, Chiriboga DA, Okazaki S. Attitudes toward mental health services: age-group differences in Korean American adults. Aging Ment Health 2009 Jan;13(1):127-134. [FREE Full text] [doi: 10.1080/13607860802591070] [Medline: 19197698]
9. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
10. Ben-Zeev D, Young MA, Madsen JW. Retrospective recall of affect in clinically depressed individuals and controls. Cogn Emot 2009 Aug;23(5):1021-1040. [doi: 10.1080/02699930802607937]
11. Wenze SJ, Miller IW. Use of ecological momentary assessment in mood disorders research. Clin Psychol Rev 2010 Aug;30(6):794-804. [doi: 10.1016/j.cpr.2010.06.007] [Medline: 20619520]
12. Parsey CM, Schmitter-Edgecombe M. Using actigraphy to predict the ecological momentary assessment of mood, fatigue, and cognition in older adulthood: mixed-methods study. JMIR Aging 2019 Jan 18;2(1):e11331. [FREE Full text] [doi: 10.2196/11331] [Medline: 31518282]
13. Acebo C, LeBourgeois MK. Actigraphy. Respir Care Clin N Am 2006 Mar;12(1):23-30, viii. [doi: 10.1016/j.rcc.2005.11.010] [Medline: 16530645]
14. Baune BT, Caliskan S, Todder D. Effects of adjunctive antidepressant therapy with quetiapine on clinical outcome, quality of sleep and daytime motor activity in patients with treatment-resistant depression. Hum Psychopharmacol 2007 Jan;22(1):1-9. [doi: 10.1002/hup.817] [Medline: 17191266]
15. Burton C, McKinstry B, Szentagotai Tatar A, Serrano-Blanco A, Pagliari C, Wolters M. Activity monitoring in patients with depression: a systematic review. J Affect Disord 2013 Feb 15;145(1):21-28. [doi: 10.1016/j.jad.2012.07.001] [Medline: 22868056]
16. Kim H, Hong S, Kwon S, Kim N. Developing geriatric nursing competency using information and communication technology: focusing on a case of depression assessment. J Korean Gerontol Nurs 2018 Dec 31;20(Suppl 1):119-126. [doi: 10.17079/jkgn.2018.20.s1.s119]
17. Crowe E, Daly M, Delaney L, Carroll S, Malone KM. The intra-day dynamics of affect, self-esteem, tiredness, and suicidality in major depression. Psychiatry Res 2019 Sep;279:98-108. [doi: 10.1016/j.psychres.2018.02.032] [Medline: 29661498]
18. Kratz AL, Fritz NE, Braley TJ, Scott EL, Foxen-Craft E, Murphy SL. Daily temporal associations between physical activity and symptoms in multiple sclerosis. Ann Behav Med 2019 Jan 1;53(1):98-108. [FREE Full text] [doi: 10.1093/abm/kay018] [Medline: 29697757]
19. Goodwin M, Haghighi M, Qu T, Akcakaya M, Erdogmus D, Intille S. Moving Towards a Real-time System for Automatically Recognizing Stereotypical Motor Movements in Individuals on the Autism Spectrum Using Wireless Accelerometry. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp'14; September 13-17, 2014; Seattle, Washington, USA p. 861-872. [doi: 10.1145/2632048.2632096]
20. Puiatti A, Mudda S, Giordano S, Mayora O. Smartphone-Centred Wearable Sensors Network for Monitoring Patients With Bipolar Disorder. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011 Presented at: EMBC'11; August 30-September 3 2011; Boston, MA, USA p. 3644-3647. [doi: 10.1109/IEMBS.2011.6090613]
21. Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. IEEE Trans Inf Technol Biomed 2009 Nov;13(6):864-873. [FREE Full text] [doi: 10.1109/TITB.2009.2033471] [Medline: 19846382]
22. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing suicide risk and emotional distress in chinese social media: a text mining and machine learning study. J Med Internet Res 2017 Jul 10;19(7):e243. [FREE Full text] [doi: 10.2196/jmir.7276] [Medline: 28694239]
23. Hung GC, Yang PC, Chang CC, Chiang JH, Chen YY. Predicting negative emotions based on mobile phone usage patterns: an exploratory study. JMIR Res Protoc 2016 Aug 10;5(3):e160. [FREE Full text] [doi: 10.2196/resprot.5551] [Medline: 27511748]
24. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111. [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
25. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 8;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
26. Berrouiguet S, Perez-Rodriguez MM, Larsen M, Baca-García E, Courtet P, Oquendo M. From eHealth to iHealth: transition to participatory and personalized medicine in mental health. J Med Internet Res 2018 Jan 3;20(1):e2. [FREE Full text] [doi: 10.2196/jmir.7412] [Medline: 29298748]
27. Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. Third edition. Waltham, MA: Elsevier; 2011.
28. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323. [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]

29. Kang Y, Na D, Hahn S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. J Clin Neurol 1997;15(2):300-308. [FREE Full text]
30. Korea Suicide Prevention Center. 2013. Suicide crisis counseling manual. URL: http://www.spckorea.or.kr/s_board/download.php?OFileName=%C0%DA%BB%EC%C0%A7%B1%E2%BB%F3%B4%E3+%B8%C5%B4%BA%BE%F3.pdf&FileName=130719054627_weblibrary_1340_4006_1.pdf [accessed 2019-02-28]
31. Eldahan AI, Pachankis JE, Jonathon Rendina H, Ventuneac A, Grov C, Parsons JT. Daily minority stress and affect among gay and bisexual men: a 30-day diary study. J Affect Disord 2016 Jan 15;190:828-835. [FREE Full text] [doi: 10.1016/j.jad.2015.10.066] [Medline: 26625095]
32. Bae JN, Cho MJ. Development of the Korean version of the Geriatric Depression Scale and its short form among elderly psychiatric patients. J Psychosom Res 2004 Sep;57(3):297-305. [doi: 10.1016/j.jpsychores.2004.01.004] [Medline: 15507257]
33. Yi JS, Bae SO, Ahn YM, Park DB, Noh KS, Shin HK. Validity and reliability of the Korean version of the Hamilton Depression Rating Scale (K-HDRS). J Korean Neuropsychiatr Assoc 2005 Jul;44(4):456-465. [FREE Full text]
34. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
35. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011 Aug;15(4):259-267. [doi: 10.1016/j.smrv.2010.10.001] [Medline: 21237680]
36. Thorne HC, Jones KH, Peters SP, Archer SN, Dijk D. Daily and seasonal variation in the spectral composition of light exposure in humans. Chronobiol Int 2009 Jul;26(5):854-866. [doi: 10.1080/07420520903044315] [Medline: 19637047]
37. Vachon H, Bourbousson M, Deschamps T, Doron J, Bulteau S, Sauvaget A, et al. Repeated self-evaluations may involve familiarization: an exploratory study related to ecological momentary assessment designs in patients with major depressive disorder. Psychiatry Res 2016 Nov 30;245:99-104. [doi: 10.1016/j.psychres.2016.08.034] [Medline: 27541343]
38. Alpaydin E. Introduction To Machine Learning. Second edition. Cambridge, MA: The MIT Press; 2004.
39. Jeon SW, Yoon SY, Ko YH, Joe SH, Kim YK, Han C, et al. Do somatic symptoms predict the severity of depression? A validation study of the Korean version of the depression and somatic symptoms scale. J Korean Med Sci 2016 Dec;31(12):2002-2009. [FREE Full text] [doi: 10.3346/jkms.2016.31.12.2002] [Medline: 27822942]
40. Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton Depression Rating Scale. J Affect Disord 2013 Sep 5;150(2):384-388. [doi: 10.1016/j.jad.2013.04.028] [Medline: 23759278]
41. Lee SC, Kim WH, Chang SM, Kim BS, Lee DW, Bae JN. The use of the Korean version of Short Form Geriatric Depression Scale (SGDS-K) in the community dwelling elderly in Korea. J Korean Geriatr Psychiatry 2013 Jun;17(1):36-42. [FREE Full text]
42. Sathyanarayana A, Joty S, Fernandez-Luque L, Ofli F, Srivastava J, Elmagarmid A, et al. Sleep quality prediction from wearable data using deep learning. JMIR Mhealth Uhealth 2016 Nov 4;4(4):e125. [FREE Full text] [doi: 10.2196/mhealth.6562] [Medline: 27815231]
43. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: Building a Mood Sensor From Smartphone Usage Patterns. In: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services. 2013 Presented at: MobiSys'13; June 25-28, 2013; Taipei, Taiwan p. 389-402. [doi: 10.1145/2462456.2464449]
44. Insel TR. Digital phenotyping: technology for a new science of behavior. J Am Med Assoc 2017 Oct 3;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
45. Tanaka T, Kokubo K, Iwasa K, Sawa K, Yamada N, Komori M. Intraday activity levels may better reflect the differences between major depressive disorder and bipolar disorder than average daily activity levels. Front Psychol 2018;9:2314. [FREE Full text] [doi: 10.3389/fpsyg.2018.02314] [Medline: 30581399]
46. Ravesloot C, Ward B, Hargrove T, Wong J, Livingston N, Torma L, et al. Why stay home? Temporal association of pain, fatigue and depression with being at home. Disabil Health J 2016 Apr;9(2):218-225. [doi: 10.1016/j.dhjo.2015.10.010] [Medline: 26750976]
47. Wassink-Vossen S, Collard RM, Oude Voshaar RC, Comijs HC, de Vocht HM, Naarding P. Physical (in)activity and depression in older people. J Affect Disord 2014 Jun;161:65-72. [doi: 10.1016/j.jad.2014.03.001] [Medline: 24751309]
48. Schuch F, Vancampfort D, Firth J, Rosenbaum S, Ward P, Reichert T, et al. Physical activity and sedentary behavior in people with major depressive disorder: a systematic review and meta-analysis. J Affect Disord 2017 Mar 1;210:139-150. [doi: 10.1016/j.jad.2016.10.050] [Medline: 28033521]
49. Seo JY, Chao Y. Effects of exercise interventions on depressive symptoms among community-dwelling older adults in the United States: a systematic review. J Gerontol Nurs 2018 Mar 1;44(3):31-38. [doi: 10.3928/00989134-20171024-01] [Medline: 29077980]
50. Cong L, Dou P, Chen D, Cai L. Depression and associated factors in the elderly cadres in Fuzhou, China: a community-based study. International Journal of Gerontology 2015;9(1):29-33. [doi: 10.1016/j.ijge.2015.02.001]
51. Ouyang P, Sun W. Depression and sleep duration: findings from middle-aged and elderly people in China. Public Health 2019 Jan;166:148-154. [doi: 10.1016/j.puhe.2018.10.007] [Medline: 30503968]

52. Triantafillou S, Saeb S, Lattie EG, Mohr DC, Kording KP. Relationship between sleep quality and mood: ecological momentary assessment study. JMIR Ment Health 2019 Mar 27;6(3):e12613. [FREE Full text] [doi: 10.2196/12613] [Medline: 30916663]
53. Kim HR. Comparison of prevalence of depression and its risk factors among the elderly female and male living alone in Korea. Korean Public Health Res 2014 Feb;40(1):13-27. [FREE Full text]
54. Yu J, Rawtaer I, Fam J, Jiang M, Feng L, Kua EH, et al. Sleep correlates of depression and anxiety in an elderly Asian population. Psychogeriatrics 2016 May;16(3):191-195. [FREE Full text] [doi: 10.1111/psyg.12138] [Medline: 26179204]
55. Kang E, Lee M. Identifying the effect of living alone on life in later adulthood: comparison between living alone and those living with others with a propensity score matching analysis. Health Soc Welf Rev 2018 Dec;38(4):196-226. [doi: 10.15709/hswr.2018.38.4.196]
56. Friberg TR, Borrero G. Diminished perception of ambient light: a symptom of clinical depression? J Affect Disord 2000 Dec;61(1-2):113-118. [doi: 10.1016/s0165-0327(99)00194-9] [Medline: 11099749]
57. Yu B, Hu J, Feijs L. Design and Evaluation of an Ambient Lighting Interface of HRV Biofeedback System in Home Setting. In: Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence. 2014 Presented at: UCAmI'14; December 2-5, 2014; Belfast, UK. [doi: 10.1007/978-3-319-13102-3_17]
58. Yu B, Hu J, Funk M, Feijs L. DeLight: biofeedback through ambient light for stress intervention and relaxation assistance. Pers Ubiquit Comput 2018 May 23;22(4):787-805. [doi: 10.1007/s00779-018-1141-6]
59. Rullier L, Atzeni T, Husky M, Bouisson J, Dartigues JF, Swendsen J, et al. Daily life functioning of community-dwelling elderly couples: an investigation of the feasibility and validity of ecological momentary assessment. Int J Methods Psychiatr Res 2014 Jun;23(2):208-216. [doi: 10.1002/mpr.1425] [Medline: 24375556]
60. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. J Technol Hum Serv 2016;34(2):209-223. [FREE Full text] [doi: 10.1080/15228835.2016.1170649] [Medline: 27683018]
61. Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. J Med Internet Res 2019 Apr 17;21(4):e11029. [FREE Full text] [doi: 10.2196/11029] [Medline: 30994461]
62. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-26. [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
63. Cain AE, Depp CA, Jeste DV. Ecological momentary assessment in aging research: a critical review. J Psychiatr Res 2009 Jul;43(11):987-996. [FREE Full text] [doi: 10.1016/j.jpsychires.2009.01.014] [Medline: 19272611]
64. Düking P, Hotho A, Holmberg HC, Fuss FK, Sperlich B. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. Front Physiol 2016;7:71. [FREE Full text] [doi: 10.3389/fphys.2016.00071] [Medline: 27014077]
65. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. JAMA Psychiatry 2019 Feb 1;76(2):190-198. [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]
66. Collins S, Sabanovic S, Fraune M, Randall N, Eldridge L, Piatt JA. Sensing Companions: Potential Clinical Uses of Robot Sensor Data for Home Care of Older Adults with Depression. In: Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. 2018 Presented at: HRI'18; March 5-8, 2018; Chicago, IL, USA p. 89-90. [doi: 10.1145/3173386.3177047]
67. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]
68. Gordon I, Ling J, Robinson L, Hayes C, Crosland A. Talking about depression during interactions with GPs: a qualitative study exploring older people's accounts of their depression narratives. BMC Fam Pract 2018 Nov 3;19(1):173. [FREE Full text] [doi: 10.1186/s12875-018-0857-8] [Medline: 30390637]

Abbreviations

AUC: area under receiver operating characteristic curve
EMA: ecological momentary assessment
K-HDRS: Korean version of the Hamilton Depression Rating Scale
MDD: major depressive disorder
NRF: National Research Foundation
RGB: red, green, and blue
SGDS-K: Korean version of the Short Geriatric Depression Scale
WASO: wake-after-sleep-onset

Edited by G Eysenbach; submitted 29.03.19; peer-reviewed by Q Cheng, SM Kia, G Norman; comments to author 24.04.19; revised version received 30.07.19; accepted 30.08.19; published 16.10.19.

Please cite as:
Kim H, Lee S, Lee S, Hong S, Kang H, Kim N
Depression Prediction by Using Ecological Momentary Assessment, Actiwatch Data, and Machine Learning: Observational Study on Older Adults Living Alone
JMIR Mhealth Uhealth 2019;7(10):e14149
URL: http://mhealth.jmir.org/2019/10/e14149/
doi:10.2196/14149
PMID:31621642

Original Paper

Exploring the Time Trend of Stress Levels While Using the Crowdsensing Mobile Health Platform, TrackYourStress, and the Influence of Perceived Stress Reactivity: Ecological Momentary Assessment Pilot Study

Rüdiger Pryss[1], PhD; Dennis John[2], Prof Dr; Winfried Schlee[3], PD, PhD; Wolff Schlotz[4], PD, PhD; Johannes Schobel[1], PhD; Robin Kraft[1], MSc; Myra Spiliopoulou[5], Prof Dr; Berthold Langguth[3], Prof Dr; Manfred Reichert[1], Prof Dr; Teresa O'Rourke[6], BSc; Henning Peters[7], MD, PhD; Christoph Pieh[6], Prof Dr; Claas Lahmann[8], Prof Dr; Thomas Probst[6], Prof Dr

[1]Institute of Databases and Information Systems, Ulm University, Ulm, Germany

[2]Lutheran University of Applied Sciences, Nuremberg, Germany

[3]Department of Psychiatry and Psychotherapy, University of Regensburg at Bezirksklinikum, Regensburg, Germany

[4]Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

[5]Faculty of Computer Science, Otto-von-Guericke-University, Magdeburg, Germany

[6]Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, Austria

[7]Department of Psychiatry and Psychotherapy, LMU Munich, Munich, Germany

[8]Faculty of Medicine, Department of Psychosomatic Medicine and Psychotherapy, Medical Center-University of Freiburg, Freiburg, Germany

Corresponding Author:
Rüdiger Pryss, PhD
Institute of Databases and Information Systems
Ulm University
James-Franck-Ring
Ulm, 89081
Germany
Phone: 49 731 502 4136
Email: ruediger.pryss@uni-ulm.de

Abstract

Background: The mobile phone app, TrackYourStress (TYS), is a new crowdsensing mobile health platform for ecological momentary assessments of perceived stress levels.

Objective: In this pilot study, we aimed to investigate the time trend of stress levels while using TYS for the entire population being studied and whether the individuals' perceived stress reactivity moderates stress level changes while using TYS.

Methods: Using TYS, stress levels were measured repeatedly with the 4-item version of the Perceived Stress Scale (PSS-4), and perceived stress reactivity was measured once with the Perceived Stress Reactivity Scale (PSRS). A total of 78 nonclinical participants, who provided 1 PSRS assessment and at least 4 repeated PSS-4 measurements, were included in this pilot study. Linear multilevel models were used to analyze the time trend of stress levels and interactions with perceived stress reactivity.

Results: Across the whole sample, stress levels did not change while using TYS (P=.83). Except for one subscale of the PSRS, interindividual differences in perceived stress reactivity did not influence the trajectories of stress levels. However, participants with higher scores on the PSRS subscale reactivity to failure showed a stronger increase of stress levels while using TYS than participants with lower scores (P=.04).

Conclusions: TYS tracks the stress levels in daily life, and most of the results showed that stress levels do not change while using TYS. Controlled trials are necessary to evaluate whether it is specifically TYS or any other influence that worsens the stress levels of participants with higher reactivity to failure.

(JMIR Mhealth Uhealth 2019;7(10):e13978) doi:10.2196/13978

KEYWORDS

mHealth; psychological stress; crowdsensing; ecological momentary assessment; pilot study

Introduction

Background

Selye introduced the term stress as the "nonspecific response of the body to any demand made upon it" [1]. According to the transactional model of stress by Lazarus and Folkman [2], 2 cognitive processes (primary and secondary appraisal) determine the individual experience of stress. Primary appraisal reflects an individual's evaluation of the situation being relevant and potential threats, whereas secondary appraisal reflects an individual's evaluation of manageability of the situation. If the situation is considered relevant and one's own capacities are considered insufficient to deal with the situation, then stress is the consequence. Note that neuroscience showed that stress involves reactions in the central and peripheral nervous system [3].

Continuously elevated stress levels can increase the risk of mental and somatic disease [3,4]. Stress levels vary between and within individuals. Individuals differ in their stress levels because of their genetics, developmental experiences, and personality traits to provide only some examples [3]. An important concept in this context constitutes perceived stress reactivity, which has been defined as *a disposition that underlies relatively stable individual differences in stress responses*. [5]. Within an individual, stress levels can vary depending on, for instance, the severity, intensity, and the frequencies of stressors [3]. Therefore, the assessment of inter- and intraindividual aspects of stress levels is of paramount importance, for example, for reducing the risk to develop mental or somatic disorders.

Objectives

Ecological momentary assessments (EMAs) of stress levels allow the investigation of these inter- and intraindividual differences under real-life conditions [6]. Although a recent review on mobile phone–based stress assessments included 35 studies [7], validated stress scales were used only in 5 of these studies. In this study, we present the *TrackYourStress* (TYS) crowdsensing mobile health (mHealth) platform that comprises the short version of the validated Perceived Stress Scale (PSS-4; [8]). Although the PSS-4 received some criticism [9,10], both the reliability and validity were acceptable in a European study [11]. We selected the PSS-4 instead of, for example, longer versions of the PSS [8] or the Perceived Stress Questionnaire (PSQ; [12]) for TYS, as scales for mobile phone–based assessments should be both psychometrically sound and as short as possible, so that participants are willing to fill in the scale continuously over a longer period of time and without (1) any bias because of measurement reactivity or (2) missing data because of missed signals. TYS combines EMAs and crowdsensing by solely using mobile technology and by integrating mobile phone sensors to collect data. In another study [13], we investigated passively sensed environmental data (Global Positioning System [GPS] location of TYS users) as predictors of daily measured stress-related items. In contrast to the previous study, this paper introduces TYS in more detail and reports results of an explorative pilot study on the trajectories of perceived stress levels assessed weekly with the PSS-4 in nonclinical TYS users to investigate measurement reactivity, that is, the extent to which measuring affects the measurement itself. The following 2 research questions (RQ) were addressed:

- RQ1 refers to average measurement reactivity in the total sample: How is the time trend of stress levels when using TYS in the entire population being studied in a first explorative analysis?
- RQ2 refers to interindividual differences in measurement reactivity: Is there an interaction between the time trend of stress levels when using TYS and interindividual differences in perceived stress reactivity? As perceived stress reactivity and stress levels have been shown to be correlated in the cross-sectional research [5], we explored whether the individuals' perceived stress reactivity interacts not only with simultaneously measured stress levels but also with longitudinally measured stress trajectories.

Methods

Measures

Perceived Stress Scale

The PSS-4 [8] was used in this study to operationalize stress levels using repeated measures across at least 4 measurement occasions. Items were scored on a Likert scale ranging from 0 to 4. In items 1 and 4, higher scores indicate more stress (eg, *…how often have you felt that you were unable to control*...), but in items 2 and 3, higher scores indicate less stress (eg, *…how often have you felt confident about*...). Therefore, items 2 and 3 had to be inverted—(0=4) (1=3) (2=2) (3=1) (4=0)—to calculate the PSS-4 scale score so that higher scores indicate higher stress levels. The PSS-4 was repeatedly assessed on the mobile phone of TYS users. The implemented instruction was to rate the PSS-4 for the last week. The intraclass correlation (ICC) for the PSS-4 scale scores was rather high at ICC=0.70, suggesting a strong between-subject variance component. Using a multilevel confirmatory factor analysis framework [14], we estimated within- and between-subject reliability by calculating 2-level composite reliability (omega), which is appropriate for unit-weighted scoring of congeneric scales [15]. This resulted in ω_{within}=0.60 (95% CI 0.51 to 0.68) and $\omega_{between}$=0.93 (95% CI 0.90 to 0.96).

Perceived Stress Reactivity Scale

The PSRS [5] was assessed once at the beginning of the study by each TYS user on the mobile phone. It measures stress reactivity, that is, interindividual differences in stress responses. The PSRS consists of 23 items, which were scored on a Likert scale ranging from 0 to 2. Higher scores indicate more perceived stress reactivity in some items (eg, *When tasks and duties build up to the extent that they are hard to manage…*), but less perceived stress reactivity in other items (eg, *When I want to relax after a hard day at work…*). Therefore, the following

items had to be inverted—(0=2) (1=1) (2=0)—to calculate the perceived stress reactivity subscales and total scale so that higher scale scores indicate higher perceived stress reactivity: Items 2, 10, 20, 8, 13, 15, 18, 5, 17, 19, 11, 22. The subscales and their internal consistencies (Cronbach's Alpha) in our sample were as follows: *prolonged reactivity* (alpha=.76), *reactivity to failure* (alpha=.63), *reactivity to social conflicts* (alpha=.74), *reactivity to work overload* (alpha=.86), and *reactivity to social evaluation* (alpha=.66). The total scale had an internal consistency (Cronbach alpha) of alpha=.87 in our sample.

TrackYourStress Mobile Health Crowdsensing Platform

TYS is an mHealth crowdsensing platform, which offers a website (registration and account management), an Android and iOS mobile app, a MariaDB relational database for the central repository to store the collected data, and a sophisticated Representational State Transfer (ie, RESTful) application program interface (API) for communication between the mobile apps, website, and database. A total of 4 questionnaire types, all related to stress, were implemented and integrated into TYS, namely registration, daily, weekly, and monthly questionnaire. In addition, the environmental sound level and the GPS position can be measured by the mobile apps, but TYS users must allow these sensor measurements when registering to TYS.

The applied procedure for all TYS users, in turn, is as follows: First, they register through the website or the mobile apps. Second, users have to fill in the registration questionnaire once. This includes demographic data (eg, gender and date of birth), the PSS-4, the PSRS, and the coping scales of the Stress and Coping Inventory (SCI; [16]). In a future version of TYS, a personality measure will be included in the registration questionnaire as well. Following the completion of the registration questionnaire the continuous mobile crowdsensing procedure starts, that is, filling out daily, weekly, and monthly questionnaires as well as automatically measuring the environmental sound level and GPS position. Note that the questions of the daily questionnaire can be obtained from Table 1. In particular, 3 basic user interface elements were implemented to answer a question by a TYS user. First, we implemented sliders, which represent the Visual Analogue Scales. Second, only for Question 5 *What stresses you at the moment?* we implemented a user interface element called Category, which, in turn, shows the following 4 categories to a TYS user: nothing, work-related matters, private matters, other. The user can then select all those categories, which are currently stressful for him or her. Third, we implemented a user interface element to provide the so-called Self-Assessment Manikins (SAM) [17] on Android and iOS. The SAM, in turn, are built on pictograms and used in psychology to measure emotions. To get a better impression of selected user interface elements, Figure 1 shows how the questionnaires are presented on the 2 mobile operating systems. Finally, note that the weekly questionnaire includes the PSS-4, whereas the monthly questionnaire includes the coping scales of the SCI.

Table 1. Items of the TrackYourStress daily questionnaire.

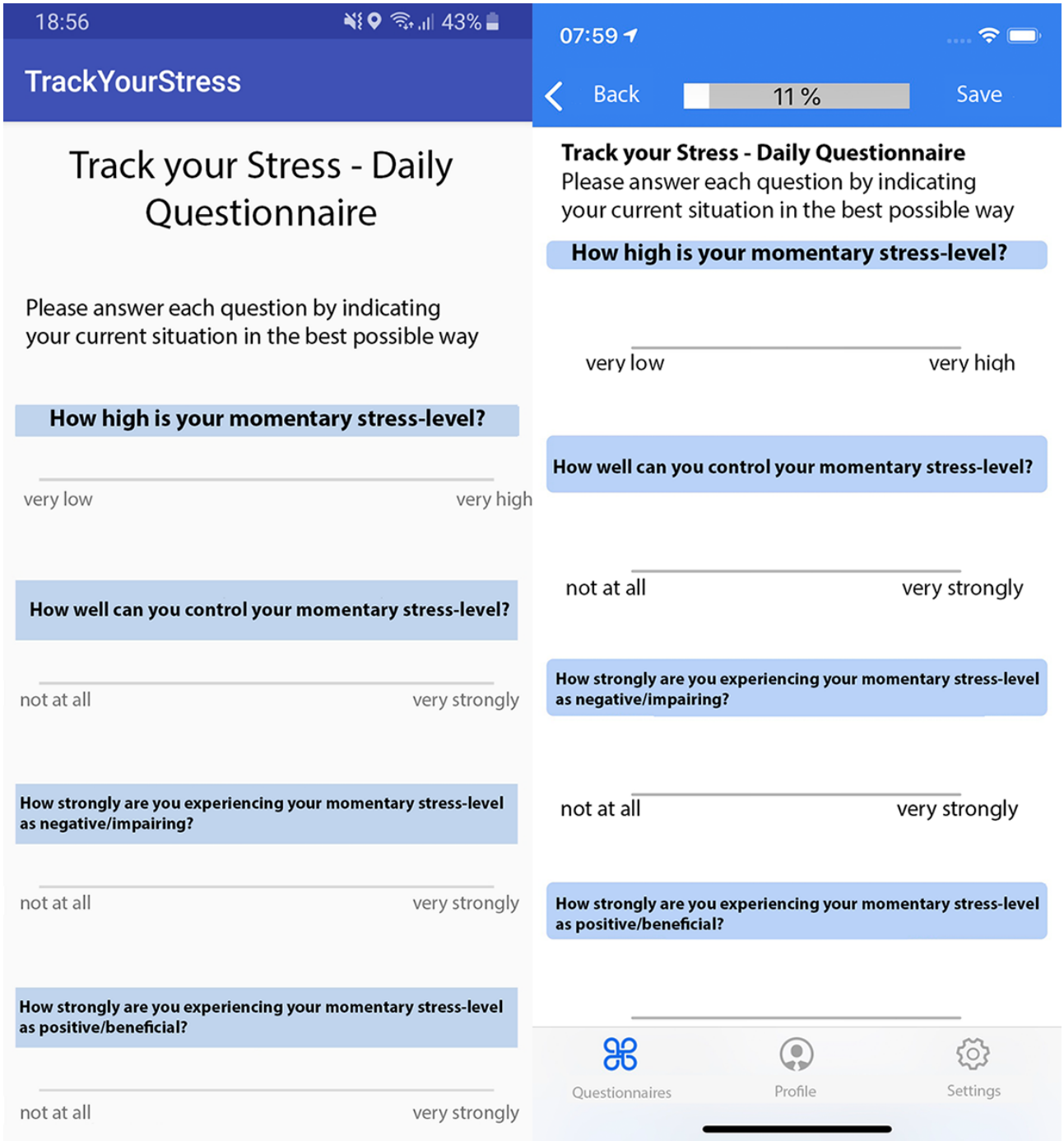
Number	Question	Scale
1	How high is your momentary stress level?	VAS[a]
2	How well can you control your momentary stress level?	VAS
3	How strongly are you experiencing your momentary stress level as negative/impairing?	VAS
4	How strongly are you experiencing your momentary stress level as positive/beneficial?	VAS
5	What stresses you at the moment?	C[b]
6	How is your mood right now?	SAM[c]
7	How is your arousal right now?	SAM
8	How important is the current situation for you personally?	VAS
9	How would you assess your ability to cope with the currently experienced situation?	VAS

[a]VAS: Visual Analogue Scale.

[b]C: categories.

[c]SAM: Self-Assessment Manikins [17].

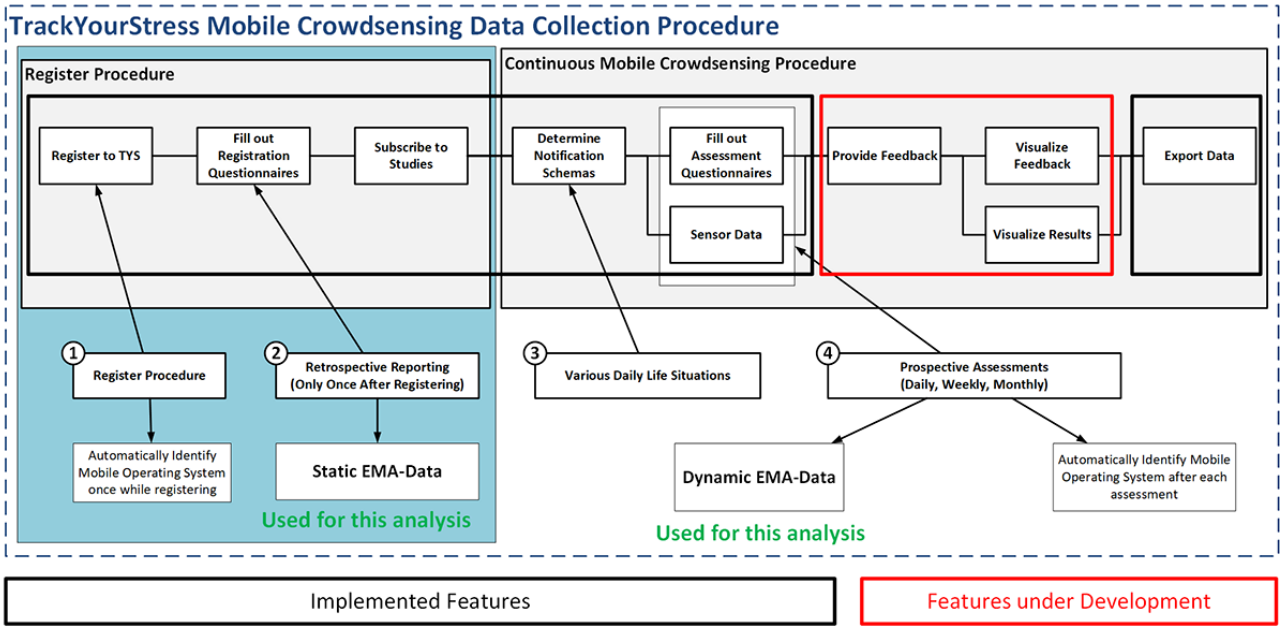
Figure 1. Impression of the daily assessment questionnaire (left Android; right iOS).



For the crowdsensing procedure, users have to accept or select a predefined notification schema. It determines the frequency and in what way (ie, fixed or random points in time) the daily, weekly, and monthly questionnaires are applied. Each time a notification appears, the user may click on it to start the mobile app (if not already running), and the respective questionnaire (daily, weekly, or monthly) is then directly shown to the user. Then, he/she can fill in the questionnaire. It is also possible for users to fill in the questionnaires without notifications. While filling out the questionnaire, either with or without using a notification, the GPS position and the environmental sound level are measured (if the app is allowed to measure them). After completion, the results are transferred to the database through the RESTful API if the mobile app is on the Web, otherwise the results are locally stored until the device gets a Web-based connection. For a detailed technical description of these features see [18,19]. The website and the RESTful API of TYS are publicly released, whereas the smart mobile apps are not yet distributed through the official mobile app stores from Apple and Google. Therefore, we used a TestFlight-based distribution for iOS and downloadable Android packages for Android. Presently, TYS is only available in German; however, we are currently translating it to English. Figure 2 summarizes the data collection procedure for TYS with its implemented and planned features. Regarding TYS in general, 2 further important aspects are finally mentioned. First, the source code of TYS is currently not freely available but will be released to the research community in the near future (ie, after several aspects have been

added for using it more conveniently; eg, by adding English as another possible language to the platform). Second, we plan to use TYS for several other purposes beyond this study. For example, it is planned to offer it to companies to anonymously track the stress levels of their employees over time. Furthermore, it is planned to give individuals the possibility to adjust the existing questionnaires to their individual needs. If they feel that other questions would fit better to their individual situation, then they should be able to adjust the daily, weekly, and monthly questionnaire. In addition, it is planned to integrate more sensors like the ones from Empatica (eg, [20]) to provide even more opportunities to the TYS users.

Figure 2. TrackYourStress mobile crowdsensing collection procedure.



Study Design

A total of 3 students of the FOM University of Applied Sciences in Augsburg and Munich (Germany) recruited participants. The 3 students asked their social networks (students, friends, family members, and colleagues) whether they are interested to partake in the study. Participants interested to partake first had to provide written informed consent before they got access to TYS. They downloaded the app and went through the TYS procedure as described earlier. The participants were informed that they should use TYS for at least 4 weeks in their daily life.

Participants

After excluding test users, N=113 individuals used TYS for this study. The participants were asked to provide at least 5 PSS-4 assessments during the study interval, that is, filling in the PSS-4 in the TYS registration questionnaire and filling in at least 4 weekly TYS questionnaires in the upcoming weeks. One weekly TYS questionnaire less than intended was tolerated to address the RQs. Therefore, the inclusion criteria for this study were filling in the PSRS and the PSS-4 in the registration questionnaire and completing at least three weekly questionnaires including the PSS-4. We deleted all PSS-4 assessments given within an interassessment interval of 24 hours, as some users filled in the PSS-4 several times a day. This resulted in a sample of 78 participants for this study. The sample description is provided in Table 2 as is the comparison between included (n=78) and excluded study participants (n=35) in baseline variables (gender, age, PSS-4 at registration, and PSRS scales at registration). Included and excluded participants differed in age, with included participants being significantly older than excluded participants (P=.005). The corresponding effect size was medium (Hedges g=0.52). No significant differences emerged for gender, stress level at registration, and perceived stress reactivity at registration.

Table 2. Sample description and statistical comparisons between included and excluded participants in baseline variables.

Variable	Included sample (n=78)	Excluded sample (n=35)	Values	
			t (*df*)	*P* value
Female, n (%)	50 (64.1)	15 (44.1)	—[a]	.06[b]
Age (years), mean (SD)	35.04 (12.30)	28.98 (9.22)	−2.90 (85.82)	.005
PSS-4[c], mean (SD)	5.32 (3.00)	5.17 (3.18)	−0.24 (111)	.81
PSRS[d] prolonged reactivity, mean (SD)	2.81 (1.97)	2.94 (1.80)	0.35 (111)	.73
PSRS reactivity to failure, mean (SD)	4.73 (1.58)	4.54 (1.63)	−0.58 (111)	.57
PSRS reactivity to social conflicts, mean (SD)	6.32 (2.21)	5.69 (2.40)	−1.37 (111)	.17
PSRS reactivity to work overload, mean (SD)	3.60 (2.65)	3.06 (2.27)	−1.06 (111)	.29
PSRS reactivity to social evaluation, mean (SD)	4.59 (2.46)	4.06 (2.51)	−1.06 (111)	.29
PSRS total, mean (SD)	22.05 (7.69)	20.29 (8.29)	−1.10 (111)	.27

[a]Not applicable.

[b]Fisher exact test.

[c]PSS: Perceived Stress Scale.

[d]PSRS: Perceived Stress Reactivity Scale.

Our average PSS-4 baseline score of 5.32 is comparable with a previously reported mean of 5.43 based on a large sample of N=37,451 participants [11]. A *t* test using the means, standard deviations, and sample sizes showed that our PSS-4 mean is not significantly different ($t_{37,527}$=−0.328; *P*=.74) from this previously reported one [11]. Our PSRS scores correspond to the ones previously depicted by Schlotz et al (see Figure 2 in their paper) [5] for a German sample between 26 and 60 years. We could not statistically compare the PSRS scores based on means, standard deviations, and sample sizes, as they are only graphically illustrated in the study by Schlotz et al [5].

Statistics

IBM SPSS v25, Stata v15.1, and Mplus v7.3 were used for statistical analyses. All statistical tests were 2-tailed, and the significance level was set to *P*<.05. To address the RQs, linear multilevel models were used. Multilevel models account for the nested data structure (assessments nested within users), are flexible in handling missing data, do not require the same amount of data points per participants, and do not require equidistant measurement points [21-23]. All multilevel models were calculated with the full maximum likelihood estimation and had assessments as level 1 and participants as level 2. As random terms, the random intercept and the random slope were included in all models, that is, the random slope term makes sure that time is included as a random effect. The unstructured variance-covariance matrix was selected in all models.

To address the RQs, the time variable was coded as follows. The time a user filled in the PSS-4 for the first time (TYS registration) was coded as 0. The further PSS-4 assessments of a user were coded as the amount of days after his/her first PSS-4 assessment, for example, the time of a PSS-4 assessment provided 8 days after his/her first PSS-4 assessment was coded as 8 and the time of a PSS-4 assessment provided 15 days after his/her first PSS-4 assessment was coded as 15.

For RQ1 (*How is the time trend of stress levels when using TYS in the entire population being studied in a first explorative analysis?*), the change of PSS-4 over time was evaluated. Therefore, 1 linear multilevel model with the PSS-4 as the dependent variable was performed, which investigated the fixed effect of time (days).

For RQ2 (*Is there an interaction between the time trend of stress levels when using TYS and interindividual differences in perceived stress reactivity?*), the interaction between time (days) and perceived stress reactivity was in focus. Thereby, 2 linear multilevel models were used. In both models, the PSS-4 was the dependent variable and the fixed effect of time (days) was evaluated. In addition, in 1 multilevel model, the fixed effects of the 5 subscales of the PSRS (time-invariant covariates) and their interactions with time (days) were investigated. In the other multilevel model, the total scale of the PSRS (time-invariant covariate) and its interaction with time (days) were evaluated as fixed effects. In both models, z-standardized PSRS scale scores were used.

Results

In total, the included n=78 participants, who completed at least four PSS-4 assessments, provided 380 PSS-4 assessments in total. On an average, they completed the PSS-4 for a mean 4.87 (SD 0.75) times. The average time interval between 2 consecutive PSS-4 assessments was mean 7.04 (SD 2.70) days. The average time interval between the first and the last PSS-4 assessment of a participant amounted to mean 29.23 (SD 6.77) days.

Results for Research Question 1

Tables 3 and 4 show the result of the linear multilevel model exploring the time trend of stress levels for the entire population being studied. It can be seen that the stress levels did not change over time when using TYS, since the fixed effect time (days) did not reach statistical significance (*P*=.83).

Table 3. Fixed effects of the linear multilevel model evaluating the time trend of perceived stress -levels (PSS-4) while using TrackYourStress.

Fixed effects	Estimate	SE	Values	
			t test (*df*)	*P* value
Intercept	5.397	0.344	15.677 (78.026)	<.001
Time	−0.003	0.013	−0.211 (65.862)	.83

Table 4. Random effects of the linear multilevel model evaluating the time trend of perceived stress -levels (PSS-4) while using TrackYourStress.

Random effects	Estimate	SE	Values	
			Wald Z test	*P* value
Var(Residual)	2.247	0.218	10.326	<.001
Var(Intercept)	7.910	1.485	5.328	<.001
Cov(Intercept; time)	−0.064	0.043	−1.480	.14
Var(Time)	0.008	0.002	3.421	.001

Results for Research Question 2

Tables 5 and 6 present the results of the linear multilevel model testing interactions between the time trend of PSS-4 stress levels and the PSRS subscales. For the PSRS subscale *reactivity to failure*, the time x PSRS interaction was statistically significant (*P*=.04). Figure 3 illustrates the time trend of stress at different levels of the *reactivity to failure* subscale and a margin plot showing confidence intervals for the slope parameter at different levels (simple slopes). It can be seen that the higher the individual's reactivity to failure, the more the stress levels increased over time while using TYS, although confidence intervals excluded zero only at relatively low and very high levels. Moreover, increases in the PSRS subscales *prolonged reactivity* (estimate=1.030; *P*=.002) and *reactivity to work overload* (estimate=0.895; *P*=.02) at baseline and registration respectively were associated with higher PSS-4 stress levels at baseline/registration.

Tables 7 and 8 show the results of the linear multilevel model that evaluated the interaction between the time trend of PSS-4 stress levels and the PSRS total scale. No significant interaction between changes of stress levels over time and the PSRS total scale emerged (*P*=.54). However, at baseline/registration, higher scores on the PSRS total scale were associated with higher PSS-4 stress levels (estimate=1.171; *P*<.001).

Table 5. Fixed effects of the linear multilevel model evaluating the time trend of perceived stress- levels (PSS-4) while using TrackYourStress including the 5 subscales of the Perceived Stress Reactivity Scale as z-standardized time-invariant covariates.

Fixed effects	Estimate	SE	Values	
			t test (*df*)	*P* value
Intercept	5.398	0.301	17.940 (78.800)	<.001
PSRS[a] prolonged reactivity	1.030	0.325	3.169 (79.275)	.002
PSRS reactivity to failure	0.058	0.339	0.172 (78.743)	.86
PSRS reactivity to social conflicts	−0.501	0.418	−1.198 (78.715)	.24
PSRS reactivity to work overload	0.895	0.382	2.342 (78.177)	.02
PSRS reactivity to social evaluation	0.393	0.368	1.069 (77.248)	.29
Time	−0.003	0.012	−0.238 (65.911)	.81
Time × PSRS prolonged reactivity	−0.004	0.014	−0.292 (73.000)	.77
Time × PSRS reactivity to failure	0.030	0.014	2.114 (64.951)	.04
Time × PSRS reactivity to social conflicts	0.009	0.017	0.502 (66.230)	.62
Time × PSRS reactivity to work overload	−0.014	0.016	−0.883 (63.165)	.38
Time × PSRS reactivity to social evaluation	−0.005	0.015	−0.333 (63.636)	.74

[a]PSRS: Perceived Stress Reactivity Scale.

Table 6. Random effects of the linear multilevel model evaluating the time trend of perceived stress- levels (PSS-4) while using TrackYourStress including the 5 subscales of the Perceived Stress Reactivity Scale as z-standardized time-invariant covariates.

Random effects	Estimate	SE	Values	
			Wald Z test	*P* value
Var(Residual)	2.248	0.218	10.325	<.001
Var(Intercept)	5.731	1.131	5.069	<.001
Cov(Intercept; time)	−0.061	0.037	−1.651	.10
Var(Time)	0.007	0.002	3.260	.001

[a]PSRS: Perceived Stress Reactivity Scale.

Figure 3. Illustration of the interaction between time (days) of perceived stress-levels (PSS-4) and the PSRS Reactivity to Failure (RtF) subscale. (A) Estimated simple slopes at 1 SD above and below mean RtF. (B) Margins and 95 % confidence interval for the time trend of stress-levels across a range of RtF scores.

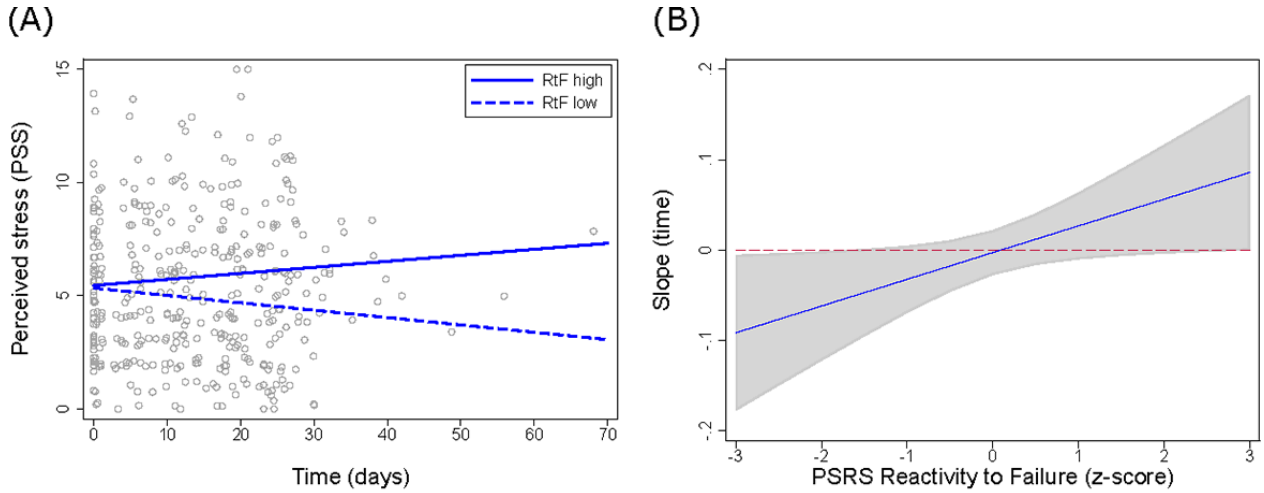


Table 7. Fixed effects of the linear multilevel model evaluating the time trend of perceived stress- levels (PSS-4) while using TrackYourStress including the total scale of the Perceived Stress. Reactivity Scale as z-standardized time-invariant covariate.

Fixed effects	Estimate	SE	Values	
			t test (*df*)	*P* value
Intercept	5.393	0.318	16.958 (78.225)	<.001
PSRS total[a]	1.171	0.320	3.658 (78.317)	<.001
Time	−0.002	0.013	−0.186(65.547)	.85
Time × PSRS total	0.008	0.013	0.613 (68.266)	.54

[a]PSRS total: Perceived Stress Reactivity Scale total score.

Table 8. Random effects of the linear multilevel model evaluating the time trend of perceived stress- levels (PSS-4) while using TrackYourStress including the total scale of the Perceived Stress Reactivity Scale as z-standardized time-invariant covariate.

Random effects	Estimate	SE	Values	
			Wald Z test	*P* value
Var(Residual)	2.252	0.219	10.302	<.001
Var(Intercept)	6.553	1.267	5.172	<.001
Cov(Intercept; time)	-0.072	0.040	−1.797	.07
Var(Time)	0.008	0.002	3.382	.001

[a]PSRS total: Perceived Stress Reactivity Scale total score.

Discussion

Principal Findings

This nonclinical study evaluated the time trend of perceived stress levels while using the crowdsensing mHealth platform, TYS, and the influence of the participants' perceived stress reactivity on longitudinally measured stress trajectories. In a first explorative analysis, we investigated the time trend of stress levels for the entire population being studied. We found no significant change of stress levels over time, and this null finding is in line with the clinical research indicating nonreactivity to EMAs [24-26]. Yet, these results should be interpreted with caution, as no change over time is the null hypothesis in statistical terms. A null hypothesis (here: no change of stress levels over time) cannot be accepted only because of a nonsignificant result; see, for example, work on equivalence and noninferiority testing [27,28]. Moreover, there are no standards on what a *true* null effect would be (eg, in terms of how narrow a confidence interval has to be), and the power of this study is far too low to test for a narrow confidence interval. In addition, we did not include a control condition and—even if there was no significant change in our study—the time trend of stress levels while using TYS could be significantly different from the time trend of stress levels in a control condition not using TYS.

Second, we analyzed whether interindividual differences in perceived stress reactivity influence the stress level trajectories while using TYS. As stress and stress reactivity were correlated in cross-sectional research [5], we wanted to explore whether the participants' perceived stress reactivity at baseline is associated with longitudinally measured stress trajectories as well. In accordance with [5], we could replicate the cross-sectional correlation between the PSRS total scale and perceived stress levels at baseline, but there were no interactions between the participants' perceived stress reactivity at baseline and the time trend of stress levels when using TYS except for one subscale of the PSRS. The higher the individual's reactivity to failure, the more the stress levels increased while using TYS. More specifically, individuals with higher perceived stress reactivity to failure (eg, mistakes during work) reported an increase in stress levels over a 4-week period. Possibly, individuals with high perceived stress reactivity to failure are more aware of failures in daily routines when monitoring their stress levels in everyday life. Being more aware of stressors in daily life might help individuals with high perceived stress reactivity to adapt their stress responses in the long run [29]. Thus, linking TYS to ecological momentary interventions (EMIs) [30] such as mobile apps for training mindfulness might be a fruitful avenue for future stress research [31-34].

Strengths and Limitations

Our research design does not allow inferring that it was specifically TYS that increased the stress levels in participants with higher reactivity to failure. There might be several confounders that influenced this result (eg, stressful life events and interpersonal problems with family or friends). This is related to the major limitation of this study, the rather low internal validity because of the lack of a control condition. A randomized controlled trial as for example planned by [35] should be performed with TYS in the future. A randomized controlled trial could compare TYS versus no TYS or specific TYS components with each other, as it could be that specific actions being executed with the TYS solution might be responsible for changes of stress levels.

Another limitation is that the reliability of the PSRS subscale *reactivity to failure* was rather low in this sample (alpha=.63). In the previous research, the Cronbach alpha values for the PSRS subscale *reactivity to failure* ranged between alpha=.65 and alpha=.73. This is likely because of the rather short scale comprising 4 items and might have biased our results. However, it should be noted that reliability estimates of that size are usually sufficient for group studies. Moreover, the external validity of the results is not very high as students of only one university recruited the sample, and the sample size was relatively small. To enhance the generalizability of the results, larger and more representative samples would be welcome. This could be realized by uploading TYS to the app stores. We plan to do this in the next months. Another threat to the external validity is the result that the TYS participants included in the current study were some years older than the TYS participants not meeting the inclusion criteria (mean 35.04 vs mean 28.98). The results might, therefore, be more representative for the TYS users being in their mid-thirties. Furthermore, the sample is rather heterogeneous, as the students recruiting the participants invited fellow students, friends, family members, colleagues, etc, to participate in the study. As we wanted to collect the data as anonymously as possible, the app did not ask about the relationship of the participants to the recruiting students (fellow student, friend, family member, and colleague). Therefore, we could not analyze how the participants' relationship to the recruiting student might have influenced the results. As we cannot find out which of our users were students, we cannot analyze whether seasonality, part of term, or time of year influences the results. This needs to be addressed in future studies. Currently, we intend to perform a study comparing stress levels of students during exam periods with stress levels of students in periods without exams. Furthermore, TYS might be a helpful tool for companies to reduce psychological health problems at the workplace. For example, TYS might help to assess stressors at the workplace and support psychosocial risk assessment. Thus, we intend to build up a large TYS database and to provide personalized feedback to users about their stress levels in relation to their reference group with employees from different companies. Users scoring higher than the reference group could be guided to EMIs (see above), internet-based self-help programs, or to professionals offering face-to-face stress management interventions nearby. It also should be kept in mind that the psychometric properties of the PSS-4 have been criticized [9,10]. Our results regarding stress levels solely rely on the self-report PSS-4. Although between-subject reliability of the outcome variable's scores was high, within-subject reliability was at a rather low level (ω_{within}=0.60). This could be because of a number of factors such as the rather high ICC of 0.70, the relatively small clusters (ie, measurements per person), or the short scale (4 items). Although this reliability might be sufficient for group studies, future versions of the app

should aim at increasing within-subject reliability of the outcome variable for achieving sufficient reliability of individual assessments. Results would be more robust when a self-report like the PSS-4 is combined with other stress assessment methods such as the Mobile Photographic Stress Meter [36] and objective measures, sensors, and computational methods [37] such as information gained from passive mobile phone sensing [38,39] or physiological signals [40,41]. Integrating such different measures of stress should be a next step in the development of TYS. Moreover, it should be investigated in a larger sample whether there are differences between iOS and Android users with regard to changes of stress levels while using TYS. Previous studies have shown that iOS and Android users differ from each other and this might confound results of mobile phone-based studies [42-44].

Conclusions

In summary, this study suggested that TYS does not change perceived stress levels in general but that TYS might influence that stress levels increase in individuals with higher reactivity to failure. These results need to be replicated in studies with a control condition and a larger more representative sample.

Acknowledgments

The authors would like to thank the students N Reineke, K Adler, and M Schönhammer for recruiting the sample.

Authors' Contributions

RP and TP substantially contributed to the TYS platform and data preparation and drafted and revised the manuscript. DJ substantially contributed to the study design, data acquisition, data interpretation, and revised the manuscript. WS and JS substantially contributed to the TYS platform and data interpretation and revised the manuscript. RK, MS, BL, MR and TO substantially contributed to the TYS platform and revised the manuscript. HP, CP, and CL substantially contributed to data interpretation and revised the manuscript.

Conflicts of Interest

None declared.

References

1. Selye H. The evolution of the stress concept. Am Sci 1973;61(6):692-699. [doi: 10.1016/0002-9149(70)90796-4] [Medline: 4746051]
2. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer Publishing Company; 1984.
3. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
4. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. J Am Med Assoc 2007 Oct 10;298(14):1685-1687. [doi: 10.1001/jama.298.14.1685] [Medline: 17925521]
5. Schlotz W, Yim IS, Zoccola PM, Jansen L, Schulz P. The perceived stress reactivity scale: measurement invariance, stability, and validity in three countries. Psychol Assess 2011 Mar;23(1):80-94. [doi: 10.1037/a0021148] [Medline: 21280954]
6. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
7. Þórarinsdóttir H, Kessing LV, Faurholt-Jepsen M. Smartphone-based self-assessment of stress in healthy adult individuals: a systematic review. J Med Internet Res 2017 Feb 13;19(2):e41. [doi: 10.2196/jmir.6397] [Medline: 28193600]
8. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
9. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127. [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
10. Ingram PB, Clarke E, Lichtenberg JW. Confirmatory factor analysis of the perceived stress scale-4 in a community sample. Stress Health 2016 Apr;32(2):173-176. [doi: 10.1002/smi.2592] [Medline: 24995556]
11. Vallejo MA, Vallejo-Slocker L, Fernández-Abascal EG, Mañanes G. Determining factors for stress perception assessed with the perceived stress scale (PSS-4) in Spanish and other European samples. Front Psychol 2018 Jan;9:37 [FREE Full text] [doi: 10.3389/fpsyg.2018.00037] [Medline: 29434563]
12. Levenstein S, Prantera C, Varvo V, Scribano ML, Berto E, Luzi C, et al. Development of the perceived stress questionnaire: a new tool for psychosomatic research. J Psychosom Res 1993 Jan;37(1):19-32. [doi: 10.1016/0022-3999(93)90120-5] [Medline: 8421257]
13. Pryss R, John D, Reichert M, Hoppenstedt B, Schmid L, Schlee W, et al. Machine Learning Findings on Geospatial Data of Users from the TrackYourStress mHealth Crowdsensing Platform. In: Proceedings of the 20th International Conference on Information Reuse and Integration for Data Science. 2019 Presented at: IRI'19; July 30-August 1, 2019; Los Angeles, CA p. 350-355 URL: http://dbis.eprints.uni-ulm.de/1812/ [doi: 10.1109/IRI.2019.00061]

14. Geldhof GJ, Preacher KJ, Zyphur MJ. Reliability estimation in a multilevel confirmatory factor analysis framework. Psychol Methods 2014 Mar;19(1):72-91. [doi: 10.1037/a0032138] [Medline: 23646988]
15. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods 2018 Sep;23(3):412-433. [doi: 10.1037/met0000144] [Medline: 28557467]
16. Satow L. Dr Lars Satow. 2012. Stress- und Coping-Inventar (SCI) URL: https://www.drsatow.de/tests/stress-und-coping-inventar/ [accessed 2019-08-03]
17. Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
18. Pryss R, Schobel J, Reichert M. Requirements for a Flexible and Generic API Enabling Mobile Crowdsensing mHealth Applications. In: Proceedings of the 4th International Workshop on Requirements Engineering for Self-Adaptive, Collaborative, and Cyber Physical Systems. 2018 Oct 22 Presented at: RESACS'18; August 20, 2018; Banff, AB, Canada p. 24-31. [doi: 10.1109/RESACS.2018.00010]
19. Pryss R, Probst T, Schlee W, Schobel J, Langguth B, Neff P, et al. Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the TrackYourTinnitus mobile platform. Int J Data Sci Anal 2018 Mar 12:-. [doi: 10.1007/s41060-018-0111-4]
20. Sevil M, Hajizadeh I, Samadi S, Feng J, Lazaro C, Frantz N, et al. Social and Competition Stress Detection With Wristband Physiological Signals. In: Proceedings of the 14th International Conference on Wearable and Implantable Body Sensor Networks. 2017 Jun 1 Presented at: BSN'17; May 9-12, 2017; Eindhoven, Netherlands p. 39-42. [doi: 10.1109/BSN.2017.7936002]
21. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford: Oxford University Press; 2003.
22. Snijders TA, Bosker RJ. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks, CA: Sage Publications; 2012.
23. Göllner R, Gollwitzer M, Heider J, Zaby A, Schröder A. [Analysis of longitudinal data with hierarchical linear models]. Zeitschrift für Klinische Psychologie und Psychotherapie 2010 Jul;39(3):179-188. [doi: 10.1026/1616-3443/a000032]
24. Schlee W, Pryss RC, Probst T, Schobel J, Bachmeier A, Reichert M, et al. Measuring the moment-to-moment variability of tinnitus: the TrackYourTinnitus smart phone app. Front Aging Neurosci 2016;8:294 [FREE Full text] [doi: 10.3389/fnagi.2016.00294] [Medline: 28018210]
25. Aaron LA, Turner JA, Mancl L, Brister H, Sawchuk CN. Electronic diary assessment of pain-related variables: is reactivity a problem? J Pain 2005 Mar;6(2):107-115. [doi: 10.1016/j.jpain.2004.11.003] [Medline: 15694877]
26. Husky M, Olié E, Guillaume S, Genty C, Swendsen J, Courtet P. Feasibility and validity of ecological momentary assessment in the investigation of suicide risk. Psychiatry Res 2014 Dec 15;220(1-2):564-570. [doi: 10.1016/j.psychres.2014.08.019] [Medline: 25155939]
27. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Soc Psychol Personal Sci 2017 May;8(4):355-362 [FREE Full text] [doi: 10.1177/1948550617697177] [Medline: 28736600]
28. Streiner DL. Unicorns do exist: a tutorial on 'proving' the null hypothesis. Can J Psychiatry 2003 Dec;48(11):756-761. [doi: 10.1177/070674370304801108] [Medline: 14733457]
29. John D, Lang FR. Subjective acceleration of time experience in everyday life across adulthood. Dev Psychol 2015 Dec;51(12):1824-1839. [doi: 10.1037/dev0000059] [Medline: 26414094]
30. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
31. Pryss R, Reichert M, John D, Frank J, Schlee W, Probst T. A Personalized Sensor Support Tool for the Training of Mindful Walking. In: Proceedings of the 15th International Conference on Wearable and Implantable Body Sensor Networks. 2018 Presented at: BSN'18; March 4-7, 2018; Las Vegas, NV, USA p. 114-117. [doi: 10.1109/bsn.2018.8329672]
32. Martín MD, de la Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mHealth: systematic search in app stores. JMIR Mhealth Uhealth 2018 May 9;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
33. Winslow BD, Chadderdon GL, Dechmerowski SJ, Jones DL, Kalkstein S, Greene JL, et al. Development and clinical evaluation of an mHealth application for stress management. Front Psychiatry 2016 Jul 26;7:130 [FREE Full text] [doi: 10.3389/fpsyt.2016.00130] [Medline: 27507949]
34. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82. [doi: 10.2196/mhealth.4328] [Medline: 26290327]
35. van Ballegooijen W, Ruwaard J, Karyotaki E, Ebert DD, Smit JH, Riper H. Reactivity to smartphone-based ecological momentary assessment of depressive symptoms (MoodMonitor): protocol of a randomised controlled trial. BMC Psychiatry 2016 Oct 21;16:359 [FREE Full text] [doi: 10.1186/s12888-016-1065-5] [Medline: 27769201]
36. Haim S, Wang R, Lord SE, Loeb L, Zhou X, Campbell AT. The Mobile Photographic Stress Meter (MPSM): A New Way to Measure Stress Using Images. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and

Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015 Presented at: UbiComp/ISWC'15 Adjunct; September 7-11, 2015; Osaka, Japan p. 733-742. [doi: 10.1145/2800835.2804398]
37. Sharma N, Gedeon T. Objective measures, sensors and computational techniques for stress recognition and classification: a survey. Comput Methods Programs Biomed 2012 Dec;108(3):1287-1301. [doi: 10.1016/j.cmpb.2012.07.003] [Medline: 22921417]
38. DaSilva AW, Huckins JF, Wang R, Wang W, Wagner DD, Campbell AT. Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e12084 [FREE Full text] [doi: 10.2196/12084] [Medline: 30888327]
39. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
40. Choi S, Kim S, Yang JS, Lee JH, Joo C, Jung HI. Real-time measurement of human salivary cortisol for the assessment of psychological stress using a smartphone. Sens Biosens Res 2014 Dec;2:8-11. [doi: 10.1016/j.sbsr.2014.08.001]
41. Mohino-Herranz I, Gil-Pita R, Ferreira J, Rosa-Zurera M, Seoane F. Assessment of mental, emotional and physical stress through analysis of physiological signals using smartphones. Sensors (Basel) 2015 Oct 8;15(10):25607-25627 [FREE Full text] [doi: 10.3390/s151025607] [Medline: 26457710]
42. Pryss R, Reichert M, Schlee W, Spiliopoulou M, Langguth B, Probst T. Differences Between Android and iOS Users of the TrackYourTinnitus Mobile Crowdsensing mHealth Platform. In: Proceedings of the 31st International Symposium on Computer-Based Medical Systems. 2018 Presented at: CBMS'18; June 18-21, 2018; Karlstad, Sweden p. 411-416. [doi: 10.1109/cbms.2018.00078]
43. Pryss R, Schlee W, Reichert M, Kurthen I, Giroud N, Jagoda L, et al. Ecological Momentary Assessment Based Differences Between Android and iOs Users of the TrackYourHearing mHealth Crowdsensing Platform. In: Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2019 Presented at: EMBC'19; July 23-27, 2019; Berlin, Germany p. 3951-3955. [doi: 10.1109/EMBC.2019.8857854]
44. Ubhi HK, Kotz D, Michie S, Schayck OC, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171 [FREE Full text] [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]

Abbreviations

API: application program interface
EMA: ecological momentary assessment
EMI: ecological momentary intervention
GPS: Global Positioning System
ICC: intraclass correlation
PSRS: Perceived Stress Reactivity Scale
PSS: Perceived Stress Scale
RESTful: Representational State Transfer
RQ: research question
SAM: Self-Assessment Manikins
TYS: TrackYourStress

Edited by G Eysenbach; submitted 11.03.19; peer-reviewed by U Scholz, J Huckins, F Seoane, J Lüscher; comments to author 27.04.19; revised version received 22.06.19; accepted 19.08.19; published 30.10.19.

Please cite as:
Pryss R, John D, Schlee W, Schlotz W, Schobel J, Kraft R, Spiliopoulou M, Langguth B, Reichert M, O'Rourke T, Peters H, Pieh C, Lahmann C, Probst T
Exploring the Time Trend of Stress Levels While Using the Crowdsensing Mobile Health Platform, TrackYourStress, and the Influence of Perceived Stress Reactivity: Ecological Momentary Assessment Pilot Study
JMIR Mhealth Uhealth 2019;7(10):e13978
URL: http://mhealth.jmir.org/2019/10/e13978/
doi:10.2196/13978
PMID:31670692

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org